



# Random differences in Szemerédi's theorem and related results

Nikos Frantzikinakis, Emmanuel Lesigne, Mate Wierdl

## ► To cite this version:

Nikos Frantzikinakis, Emmanuel Lesigne, Mate Wierdl. Random differences in Szemerédi's theorem and related results. 2013. hal-00842122v2

**HAL Id: hal-00842122**

**<https://hal.science/hal-00842122v2>**

Preprint submitted on 6 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# RANDOM DIFFERENCES IN SZEMERÉDI'S THEOREM AND RELATED RESULTS

NIKOS FRANTZIKINAKIS, EMMANUEL LESIGNE, AND MÁTÉ WIERDL

**ABSTRACT.** We introduce a new, elementary method for studying random differences in arithmetic progressions and convergence phenomena along random sequences of integers. We apply our method to obtain significant improvements on two results. The first improvement is the following: Let  $\ell$  be a positive integer and let  $u_1 \geq u_2 \geq \dots$  be a decreasing sequence of probabilities satisfying  $u_n \cdot n^{1/(\ell+1)} \rightarrow \infty$ . Let  $R = R^\omega$  be the random sequence we get by selecting the natural number  $n$  with probability  $u_n$ . If  $A$  is a set of natural numbers with positive upper density, then  $A$  contains an arithmetic progression  $a, a+r, a+2r, \dots, a+\ell r$  of length  $\ell+1$  with difference  $r \in R^\omega$ . The best earlier result (by M. Christ and us) was the condition  $u_n \cdot n^{2^{-\ell+1}} \rightarrow \infty$  with a logarithmic speed. The new bound is better when  $\ell \geq 4$ .

The other improvement concerns almost everywhere convergence of double ergodic averages: we (randomly) construct a sequence  $r_1 < r_2 < \dots$  of positive integers so that for any  $\epsilon > 0$  we have  $r_n/n^{2-\epsilon} \rightarrow \infty$  and for any measure preserving transformation  $T$  of a probability space the averages

$$\frac{1}{N} \sum_{n < N} T^n F_1(x) T^{r_n} F_2(x)$$

converge for almost every  $x$ . Our earlier best result was the  $r_n/n^{(1+1/14)-\epsilon} \rightarrow \infty$  growth rate on the sequence  $(r_n)$ .

We dedicate this paper to the 70th birthday of our friend, Karl Petersen, who raised the interest of the senior authors in random sequences 20 years ago.

## CONTENTS

1. Conventions	2
2. Introduction, History	2
2.1. Intersective and recurrent sets	2
2.2. Convergence of ergodic averages	4
2.3. Semirandom averages	5
3. General ideas behind the proofs	6
4. Proof of Theorem A, the random intersectivity theorem	8
4.1. The $\ell = 2$ case	8
4.2. The general $\ell$ case	13
5. Proof of Theorem B, the random averaging theorem	14
5.1. The $\ell = 2$ case	14
5.2. The general $\ell$ case	16
5.3. Pointwise convergence	16
6. Proof of Theorem C, the semirandom convergence theorem	21
7. Notes	26
7.1. Finitary version of Theorem A	26
7.2. Our method for bounded functions	26
7.3. Convexity methods	27
7.4. Extension to commuting transformations	28
Appendix A. Strong law of large numbers	28
References	29

---

The first author has been partially supported by Marie Curie IRG 248008.

The third author has been partially supported by the US National Science Foundation, award number DMS-1102634.

## 1. CONVENTIONS

In this paper, a *dynamical system* is a quadruple  $(X, \Sigma, \mu, T)$  where  $(X, \Sigma, \mu)$  is a probability space, and  $T$  is a  $\Sigma$ -measurable, measure preserving transformation, i.e. for any  $A \in \Sigma$ , we have  $\mu(T^{-1}A) = \mu(A)$ . Since the  $\sigma$ -algebra  $\Sigma$  doesn't play any special role in what follows, we omit it from our notation, and we just write  $(X, \mu, T)$  instead of  $(X, \Sigma, \mu, T)$ .

When we talk about functions defined on some measure space, we always assume, the function is measurable.

We use the usual notational convention that the isometry of  $L^1$  induced by a measure preserving transformation  $T$  is also denoted by the same letter  $T$ :  $Tf(x) = f(Tx)$ .

The arbitrary constants  $c$  or  $C$  appearing in an estimate are always positive, and are absolute unless otherwise noted. The same constant may have a different value even in the same series of estimates.

The notation  $\sum_{n < N}$  means  $\sum_{0 \leq n < N}$ , so 0 is included in the range of summation.

A *lacunary sequence* of positive real numbers or integers is a sequence  $\{a_1 < a_2 < a_3 < \dots < a_n < \dots\}$  such that  $\liminf_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} > 1$ .

The *cardinality* of a set  $A$  is denoted by  $|A|$ .

## 2. INTRODUCTION, HISTORY

**2.1. Intersective and recurrent sets.** For a set  $A$  of positive integers, let  $A(M)$  denote its *counting function*, the number of elements in  $A$  that are less than  $M$ ,

$$A(M) = \left| \{a \in A, a < M\} \right|.$$

We denote by  $\overline{d}(A)$  the *upper density* of  $A$

$$\overline{d}(A) = \limsup_{M \rightarrow \infty} \frac{A(M)}{M}.$$

The celebrated result of [Sze75a, Sze75b] says that if  $A$  is a set of positive integers with positive upper density, then for any positive  $\ell$  one can find an  $\ell + 1$  long arithmetic progression  $a, a + r, a + 2r, \dots, a + \ell r$  in  $A$ . One way to generalize this result is to demand that the difference  $r$  is from some prescribed sequence.

**Definition 2.1.** Let  $\ell$  be a positive integer. We say that a set  $R$  of positive integers is  $\ell$ -*recurrent* or  $\ell$ -*intersective* if any set  $A$  of positive upper density contains an  $\ell + 1$  long arithmetic progression  $a, a + r, a + 2r, \dots, a + \ell r$  with the difference  $r \in R$ .

We refer to  $\ell$  as the *order* of intersectivity.

We note that an  $\ell$ -intersective set may not be  $(\ell + 1)$ -intersective, see [FLW06].

The explanation for the terminology “intersective” is that saying  $R$  is  $\ell$ -intersective is equivalent with

$$(2.1) \quad \overline{d}(A) > 0 \text{ implies } A \cap (A - r) \cap (A - 2r) \cap \dots \cap (A - \ell r) \neq \emptyset \text{ for some } r \in R.$$

For the term “recurrent”, it is known that eq. (2.1) is equivalent with the following: In every dynamical system  $(X, \mu, T)$

$$\mu(A) > 0 \text{ implies } \mu\left(A \cap T^{-r}A \cap T^{-2r}A \cap \dots \cap T^{-\ell r}A\right) > 0 \text{ for some } r \in R.$$

The direction “recurrent  $\implies$  intersective” is known as Furstenberg's transfer principle, and it was proved in [Fur77].

It is known that the following sets are  $\ell$ -intersective

- the squares  $\{1^2, 2^2, 3^2, \dots, n^2, \dots\}$  or, more generally, the  $k$ th powers  $\{1^k, 2^k, 3^k, \dots, n^k, \dots\}$ . For  $\ell = 1$ , this was a conjecture of Lovász, and was answered in [Sár78a, Sár78b]. The case of any  $\ell$  was proved in [BL96].
- the “prime minus ones”  $\{2 - 1, 3 - 1, 5 - 1, \dots, p - 1, \dots\}$ , or the “prime plus ones”  $\{2 + 1, 3 + 1, 5 + 1, \dots, p + 1, \dots\}$  where  $p$  is a prime number. For  $\ell = 1$  this was a question raised by Erdős, and was answered by [Sár78b]. The  $\ell = 2$  case is solved in [FHK07], the case of all other  $\ell$  is solved in [WZ12].
- $\{[1^{3/2}], [2^{3/2}], [3^{3/2}], \dots, [n^{3/2}], \dots\}$  or, more generally,  $\{[1^\delta], [2^\delta], [3^\delta], \dots, [n^\delta], \dots\}$  for any fixed positive, non-integer  $\delta$ . For  $\ell = 1$ , this probably was first noted in [Wie89], but it's clear that Boshernitzan had been aware of the result. See [Bos83]. It also follows from the general result in [ES78]. For all  $\ell$ , it is proved in [FW09].

These intersective sequences don't increase faster to infinity than some power of  $n$ . In terms of the counting function  $R(N)$  this means that any of the above sets is not very thin since it satisfies  $R(N) > N^\epsilon$  for some positive  $\epsilon$ . Are there any intersective sets which increase to infinity faster than any power of  $n$ ? The question is not exactly a good one: [ES78] considered sets with dilations. A  $k$ -long dilation is a set of integers of the form  $\{d, 2d, 3d, \dots, kd\}$ . Now, if a set contains arbitrarily long dilations, then it is a consequence of Szemerédi's theorem, that the set is  $\ell$ -intersective for all  $\ell$ . This way, one can make  $\ell$ -intersective sequences with arbitrarily thin counting functions. On the other hand, a lacunary sequence cannot contain arbitrarily long dilations and it is shown in [ES77] that a lacunary sequence is not even 1-intersective.

Are then the previously mentioned intersective sets exceptional in some sense, or a typical set is also intersective? To be able to talk more conveniently about these questions, we introduce the concept of a *random sequence*:

**Definition 2.2.** Given a sequence  $\mathfrak{u}_1, \mathfrak{u}_2, \dots$  of probabilities, let  $U_1, U_2, \dots$  be a sequence of independent 0–1 valued random variables on a probability space  $(\Omega, P)$  so that  $P(U_n = 1) = \mathfrak{u}_n = 1 - P(U_n = 0)$ . We think of the sequence  $U_1(\omega), U_2(\omega), \dots$  as the indicator of the set  $R^\omega$  defined by

$$R^\omega = \{n \mid U_n(\omega) = 1\}.$$

We say that *the random sequence has some given property* if there is a set  $\Omega' \subset \Omega$  with  $P(\Omega') = 1$  so that  $R^\omega$  has the property for all  $\omega \in \Omega'$ . Note that the concept of a random sequence is always tied to a given sequence of probabilities  $\mathfrak{u}_1, \mathfrak{u}_2, \dots$ .

**WARNING 2.3.** Unless we say otherwise, in this paper we assume that the probabilities  $\mathfrak{u}_n$  form a *non-increasing* sequence,  $\{\mathfrak{u}_1 \geq \mathfrak{u}_2 \geq \mathfrak{u}_3 \geq \dots\}$ , and that they sum to  $\infty$ ,  $\sum_n \mathfrak{u}_n = \infty$ .

The standing assumption  $\sum_n \mathfrak{u}_n = \infty$  guarantees, by the strong law of large numbers (cf. Theorem A.1 in the Appendix), that the random sequence  $R$  is infinite, since its counting function  $R(N) = \sum_{n \leq N} U_n$  is almost surely asymptotically equal to  $\sum_{n \leq N} \mathfrak{u}_n$ . We make the decreasing assumption because intuitively it gives the expected behavior from a random set.

Now we can make the concept of a typical set completely precise: it is the random set associated with a given sequence of probabilities  $\mathfrak{u}_n$ . It is easy to see that if  $\mathfrak{u}_n$  goes to 0 slower than any negative power of  $n$ , that is  $\mathfrak{u}_n \cdot n^\epsilon \rightarrow \infty$  for any  $\epsilon > 0$ , then the random sequence contains arbitrarily long dilations (and intervals!), hence it is  $\ell$ -intersective for all  $\ell$ . On the other hand, if  $\mathfrak{u}_n < n^{-\epsilon}$  for some positive  $\epsilon$ , then the random sequence doesn't have arbitrarily long dilations. Nevertheless, [ES78] showed that if  $\mathfrak{u}_n > \log n \cdot n^{-1/3}$ , then the random sequence is 1-intersective. Can a random sequence be arbitrarily thin but still intersective? Well, it is shown in [Bos83] that if  $\mathfrak{u}_n \cdot n \rightarrow \infty$ , then the random sequence is still 1-intersective. Paul Balister pointed out to us that if  $\mathfrak{u}_n = \frac{1}{n(\log n)^{1/2+\epsilon}}$ , then the random sequence is lacunary and hence is not 1-intersective. Anthony Quas told us in private conversation that the method of [Bou87] can be used to show that if  $\mathfrak{u}_n = 1/n$ , then the random sequence is not 1-intersective.

While quite a lot is known about 1-intersective sets, much less is known about typical 2 or higher order intersective sets. We proved in [FLW12] that if  $\mathfrak{u}_n = n^{-b}$  for some  $b < 1/2$ , then the random set is 2-intersective.

For general  $\ell$ , we proved in [FLW12] that if

$$(2.2) \quad \mathfrak{u}_n = n^{-b} \text{ for some } 0 < b < \frac{1}{2^{\ell-1}},$$

then the random set is  $\ell$ -intersective. This result was proved independently from us in [Chr11].

The main result on random intersective sets in this current paper is the following improvement of the above results

**Theorem A** (Random  $\ell$ -intersectivity theorem). *Let  $\ell$  be a positive integer, and assume the decreasing sequence  $\mathfrak{u}_1 \geq \mathfrak{u}_2 \geq \dots$  of probabilities satisfy*

$$(2.3) \quad \mathfrak{u}_n \cdot n^{1/(\ell+1)} \rightarrow \infty.$$

*Then the random sequence is  $\ell$ -intersective.*

The condition in eq. (2.3) is better than the condition in eq. (2.2) for  $\ell \geq 4$ . The earlier result that used the condition in eq. (2.2) was obtained by using Van der Corput's inequality  $\ell - 1$  times, and this iteration poses the restriction in eq. (2.2). Our method doesn't use iteration, and it doesn't reduce the case for  $\ell$  to the

case for  $\ell - 1$ . In particular, we give a new, quite elementary proof when  $\ell = 1$ . If  $\ell = 1$ , eq. (2.3) gives the restriction  $b < 1/2$ , while we know, after Boshernitzan, that every  $b < 1$  works. Our proof doesn't use Fourier analysis or the spectral theorem.

*Remark 2.4.* We will see that instead of the individual behavior of the  $\mathfrak{u}$  probabilities, as expressed in eq. (2.3), only their average behavior is important. The exact assumption we use in the proof is

$$(2.4) \quad \frac{\sum_{n < N} \mathfrak{u}_n}{N^{1-1/(\ell+1)}} \rightarrow \infty.$$

In other words, only the counting function  $R(N)$  of the random set  $R$  matters: by the strong law of large numbers, the above is the same as saying that  $R(N)/N^{1-1/(\ell+1)} \rightarrow \infty$ .

This indicates that, in fact, the proof applies equally well to a finitary version of Theorem A; see Section 7.1.

Now that we are close to the end of this introductory section, we state a conjecture which, if true, would be an extension of Boshernitzan's result for random 1-intersective sets to  $\ell$ -intersective ones

**Conjecture 2.5.** Suppose

$$\mathfrak{u}_n \cdot n \rightarrow \infty.$$

Then the random sequence is  $\ell$ -intersective.

As we mentioned earlier, if this conjecture is true, it's sharp in the sense that if  $\mathfrak{u}_n = 1/n$ , then the random sequence is not 1-intersective.

**2.2. Convergence of ergodic averages.** In this section  $(X, \mu, T)$  denotes a dynamical system.

**Definition 2.6.** Let  $\ell$  be a positive integer. We say that the set  $R = \{r_1 < r_2 < r_3 < \dots\}$  of positive integers is  $\ell$ -averaging in norm if in any dynamical system the averages

$$\frac{1}{N} \sum_{n < N} T^{r_n} F_1 \cdot T^{2r_n} F_2 \cdots T^{\ell r_n} F_\ell$$

converge in  $L^1$  norm for any choice of bounded functions  $F_i$  on  $X$ .

While 1-averaging sequences have been studied for a long time, the study of  $\ell$ -averaging sequences for  $\ell > 1$  is more recent than that of  $\ell$ -intersective sets: That the natural numbers are 2-averaging was proved in [Fur77], the  $\ell = 3$  case was proved in [CL84], [HK01], and the breakthrough result of [HK05b], where they show that the set of natural numbers is  $\ell$ -averaging in norm, came 30 years after Szemerédi's and Furstenberg's result. Now the study of intersective and averaging sequences have been done parallel to each other: The following sets have been found  $\ell$ -averaging in norm

- the squares  $\{1^2, 2^2, 3^2, \dots, n^2, \dots\}$  or, more generally,  $\{p(1), p(2), p(3), \dots, p(n), \dots\}$ , where  $p(x)$  is polynomial taking on positive integer values for large enough  $n$ . For  $\ell = 1$ , this is a folklore result, and follows from Weyl's estimate on trigonometric sums. For all  $\ell$ , it was proved in [HK05a] (weak convergence) and in [Lei05] for norm convergence.
- $\{2, 3, 5, \dots, p, \dots\}$ , where  $p$  is a prime number. For  $\ell = 1$  this follows from Vinogradov's classical estimates on the Fourier transform (trigonometric sum) of the sequence of primes. The  $\ell = 2$  case is solved in [FHK07], the case of all other  $\ell$  is solved in [WZ12].
- $\{[1^{3/2}], [2^{3/2}], [3^{3/2}], \dots, [n^{3/2}], \dots\}$  or, more generally,  $\{[1^\delta], [2^\delta], [3^\delta], \dots, [n^\delta], \dots\}$  for any fixed positive, non-integer  $\delta$ . For  $\ell = 1$ , this may have been first noted explicitly only in [Wie89], but it's clear that Boshernitzan had been aware of the result. See [Bos83]. For all  $\ell$ , the result is proved in [Fra10].

An important difference between intersective and averaging sets is that just because a set contains arbitrarily long dilations, it may not be an averaging set. In particular, it is not easy to construct fast growing averaging sequences. To avoid trivial examples, such as sequences consisting of longer and longer intervals of consecutive integers, we are asking the following: are there any averaging sets  $\{r_1, r_2 < r_3 < \dots\}$  so that the gaps  $r_{n+1} - r_n$  between consecutive terms goes to infinity faster than any polynomial? It follows from the estimates in [Kar71] (see also [BP98, Theorem 2]), that if, for some positive  $\epsilon$ ,  $r_n = \lfloor \exp((\log n)^{3/2-\epsilon}) \rfloor$ , then  $\{r_1 < r_2 < r_3 < \dots\}$  is 1-averaging. It is not known if this sequence is 2-averaging (or, for that matter, if it's 2-intersective). How fast can an averaging sequence increase? Can it increase arbitrarily fast? It was [Bos83] (and, explicitly, [Bou88]) who showed that if

$$\mathfrak{u}_n \cdot n \rightarrow \infty,$$

then the random sequence is 1-averaging. As in case of 1-intersective sets, it is known that this result is best possible: [JLW99] proved that if

$$\mathfrak{u}_n = \frac{1}{n},$$

then the random sequence is *not* 1-averaging. How about 2-averaging random sets? Well, we still don't know of any 2-averaging random set  $\{r_1, r_2 < r_3 < \dots\}$  which goes to infinity as fast as the squares, that is when  $\mathfrak{u}_n = n^{-1/2}$ .

For all  $\ell$ , we proved in [FLW12] that if

$$\mathfrak{u}_n = n^{-b} \text{ for some } 0 < b < \frac{1}{2^{\ell-1}},$$

then the random set is  $\ell$ -averaging. This result was also proved independently from us by [Chr11]. Our main result on random  $\ell$ -averaging sets in the present paper is the following

**Theorem B** (Random  $\ell$ -averaging theorem). *Let  $\ell$  be a positive integer, and suppose the decreasing sequence  $\mathfrak{u}_1 \geq \mathfrak{u}_2 \geq \dots$  of probabilities satisfy*

$$(2.5) \quad \mathfrak{u}_n \cdot n^{1/(\ell+1)} \rightarrow \infty.$$

*Then the random sequence  $R^\omega = \{r_1 < r_2 < r_3 < \dots\}$  is  $\ell$ -averaging in norm, and the limit of the averages*

$$\frac{1}{N} \sum_{n < N} T^{r_n} F_1 \cdot T^{2r_n} F_2 \cdots T^{\ell r_n} F_\ell$$

*along the sequence is the same as the limit of the multiple averages*

$$\frac{1}{N} \sum_{n < N} T^n F_1 \cdot T^{2n} F_2 \cdots T^{\ell n} F_\ell.$$

Here, again, the real assumption the proof uses is in terms of the average behavior of the  $\mathfrak{u}_n$ , that is, in terms of the counting function of the random sequence: the assumption in eq. (2.5) can be weakened to

$$\frac{\sum_{n < N} \mathfrak{u}_n}{N^{1-1/(\ell+1)}} \rightarrow \infty.$$

We have a conjecture similar to the one for intersective sets

**Conjecture 2.7.** Suppose

$$\mathfrak{u}_n \cdot n \rightarrow \infty.$$

Then the random sequence is  $\ell$ -averaging.

In Section 5.3, we examine almost everywhere convergence, and show in Theorem D that a.e. convergence of the *difference* between the random averages and their expectations follows if we add a mild speed condition to the assumption in eq. (2.5).

**2.3. Semirandom averages.** Our third main topic is the convergence of double ergodic averages with different rates. Combinatorial applications and norm-convergence are also of interest, but the novelty is the pointwise convergence result, since it's the first result for *non-linear* sequences. For linear sequences, it was [Bou90] who proved that the averages  $\frac{1}{N} \sum_{n < N} T^n F_1 T^{2n} F_2$  converge a.e.

Our main result for “semirandom” averages is

**Theorem C** (Semirandom pointwise averaging theorem). *Suppose the decreasing sequence  $\mathfrak{u}_1 \geq \mathfrak{u}_2 \geq \dots$  of probabilities satisfy*

$$(2.6) \quad \mathfrak{u}_n \cdot \frac{n^{1/2}}{\log^{3+\delta} n} \rightarrow \infty \text{ for some } \delta > 0,$$

and

$$(2.7) \quad \mathfrak{u}_n \rightarrow 0.$$

*Then the random sequence  $R^\omega = \{r_1 < r_2 < r_3 < \dots\}$  satisfies, in every dynamical system  $(X, \mu, T)$ , for bounded functions  $F_i$ , that*

$$(2.8) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} T^n F_1(x) T^{r_n} F_2(x) = \bar{F}_1 \bar{F}_2 \text{ for a.e. } x \in X,$$

where  $\bar{F}_i$  is the  $L^2$  projection of  $F_i$  to the  $T$ -invariant functions.

**Remark 2.8.** It will be clear from the proof that if we weaken the assumption in eq. (2.6) to

$$\mathfrak{u}_n \cdot n^{1/2} \rightarrow \infty,$$

we can still conclude mean convergence, and we have recurrence (intersectivity). What we mean by intersectivity in this context is that if the set of positive integers  $A$  has positive upper density, then for some  $n$  and  $a$ , the three numbers  $a, a+n, a+r_n$  belong to  $A$ . In fact, the lower density of such  $n$ 's is at least  $(\bar{d}(A))^3$

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} \bar{d}(A \cap (A-n) \cap (A-r_n)) \geq (\bar{d}(A))^3.$$

In [FLW12], we proved eq. (2.8) with the condition  $\mathfrak{u}_n = n^{-b}$  for some  $b < 1/14$ , so with our theorem, we improve the range of  $b$  to  $b < 1/2$ .

Note that the assumption in eq. (2.7) is equivalent with the random sequence having 0 density. This assumption is *necessary* to have the limit equal  $\bar{F}_1 \bar{F}_2$ . If we drop the assumption that  $\mathfrak{u}_n \rightarrow 0$ , then we don't know if we have convergence—such as in the most basic “coin flipping” case,

**Unsolved problem 2.9.** Suppose  $\mathfrak{u}_n = 1/2$ .

Is it true that the random sequence  $R^\omega = \{r_1 < r_2 < r_3 < \dots\}$  satisfies, in every dynamical system and for bounded functions  $F_i$ , that the averages

$$\frac{1}{N} \sum_{n < N} T^n F_1(x) T^{r_n} F_2(x)$$

converge for a.e.  $x \in X$ ?

**Conjecture 2.10.** Suppose the decreasing sequence  $\mathfrak{u}_1 \geq \mathfrak{u}_2 \geq \dots$  of probabilities satisfy

$$\mathfrak{u}_n \cdot \frac{n}{(\log \log n)^{1+\delta}} \rightarrow \infty \text{ for some } \delta > 0,$$

and

$$\mathfrak{u}_n \rightarrow 0.$$

Then the random sequence  $R^\omega = \{r_1 < r_2 < r_3 < \dots\}$  satisfies, in every dynamical system, for bounded functions  $F_i$ , that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} T^n F_1(x) T^{r_n} F_2(x) = \bar{F}_1 \bar{F}_2 \text{ for a.e. } x \in X,$$

where  $\bar{F}_i$  is the projection of the function  $F_i$  to the  $T$ -invariant functions.

If  $F_1$  is  $T$ -invariant, then the conjecture is true. See [Bou88]. Even in this case, a speed of a power of  $\log \log n$  is necessary as shown in [JLW99].

### 3. GENERAL IDEAS BEHIND THE PROOFS

Our proof of Theorem A will be presented in an elementary way, but here we explain how the idea of the proof came about because we believe this is a good general context to search for further improvements.

In case of 1- and 2-intersectivity, the main tasks to prove are

$$\lim_{N \rightarrow \infty} \sup_A \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \bar{d}(A \cap (A - n)) \right| = 0 \quad \text{with probability 1,}$$

and

$$\lim_{N \rightarrow \infty} \sup_A \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \bar{d}(A \cap (A - n) \cap (A - 2n)) \right| = 0 \quad \text{with probability 1,}$$

respectively, where in  $\sup_A$  we take the supremum over all subsets of the positive integers. For 1- and 2-averaging the main tasks are

$$\lim_{N \rightarrow \infty} \sup_{(X, T, \mu), F} \int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) T^n F \right| = 0 \quad \text{with probability 1,}$$

and

$$\lim_{N \rightarrow \infty} \sup_{(X, T, \mu), F_i} \int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) T^n F_1 \cdot T^{2n} F_2 \right| = 0 \quad \text{with probability 1,}$$

respectively, where in  $\sup_{(X, T, \mu), F_i}$  we take the supremum over all dynamical systems  $(X, T, \mu)$  and indicators  $F_i$ . After a duality consideration, all these cases are proved via estimates of the form

$$(3.1) \quad P \left( \sup_{(a_1, a_2, a_3, \dots, a_{N-1}) \in \mathcal{A}} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot a_n \right| > \epsilon \right) < \exp \left( -c\epsilon^2 \sum_{n < N} \mathbb{1}_n \right),$$

where in  $\sup_{(a_1, a_2, a_3, \dots, a_{N-1}) \in \mathcal{A}}$  we take the the supremum over a certain class  $\mathcal{A}$  of vectors  $(a_1, a_2, a_3, \dots, a_{N-1})$  of real numbers with  $|a_n| \leq 1$ . For example, in case of 2-averaging, we have

$$a_n = \int_X G \cdot T^n F_1 \cdot T^{2n} F_2$$

and the supremum is taken over all dynamical systems  $(X, T, \mu)$ , indicators  $F_i$  and  $\pm 1$ -valued functions  $G$ . The inequality in eq. (3.1) for a *fixed* vector  $(a_1, a_2, a_3, \dots, a_{N-1})$  with  $|a_n| \leq 1$  follows from Bernstein's exponential inequality, Lemma 4.1 below, so the difficulty is to handle the supremum. Now, for estimating the supremum of a class of random variables  $Y_k$ ,  $k \in \mathcal{K}$ , the best estimate we have is the union estimate,

$$P \left( \sup_{k \in \mathcal{K}} Y_k > \epsilon \right) \leq |\mathcal{K}| \cdot \sup_{k \in \mathcal{K}} P(Y_k > \epsilon).$$

In our case,  $\mathcal{K}$  would be the class of vectors  $\mathcal{A} = (a_1, a_2, a_3, \dots, a_{N-1})$ . This requires the reduction of the number of vectors  $(a_1, a_2, a_3, \dots, a_{N-1})$  in the supremum: we need to be able to find a subclass  $\mathcal{A}_0$  of  $\mathcal{A}$  which has the following properties

**$\mathcal{A}_0$  approximates  $\mathcal{A}$  within  $\epsilon$ :** for any  $(a_1, a_2, a_3, \dots, a_{N-1}) \in \mathcal{A}$  there is  $(a'_1, a'_2, a'_3, \dots, a'_{N-1}) \in \mathcal{A}_0$  so that  $|a_n - a'_n| < \epsilon$ ,  $n = 1, 2, 3, \dots, N$ .

**The cardinality of  $\mathcal{A}_0$  is  $\exp(o(\sum_{n < N} \mathbb{1}_n))$ :** More precisely, for every  $\epsilon$ , we need to have  $\log(|\mathcal{A}_0|) = O(\epsilon^2 \sum_{n < N} \mathbb{1}_n)$ .

In other words, the existence of  $\mathcal{A}_0$  with the above properties means that the  $\epsilon$ -covering number of  $\mathcal{A}$  in the  $\ell^\infty$  metric needs to be  $\exp(O(\epsilon^2 \sum_{n < N} \mathbb{1}_n))$ . In both the intersective and averaging case,  $\mathcal{A}$  is uncountably infinite, but via appropriate (and more or less standard) transference arguments, we will show that in both cases  $\mathcal{A}$  has finite  $\epsilon$ -covering numbers. In case of 1-intersectivity and 1-averaging, the set  $\mathcal{A}$  contains vectors  $(a_1, a_2, a_3, \dots, a_{N-1})$  given by single correlation, that is, by convolution,

$$a_n = \frac{1}{Q} \sum_{a < Q} g(a) f(a + n),$$

where  $Q < N$ , and  $f$  and  $g$  are two-valued functions defined on the interval  $[0, 2N]$ . In this case, Parseval's formula easily yields that the  $\epsilon$ -covering number is  $O(N) = \exp(O(\log N))$  which implies that  $\sum_{n < N} \mathbb{1}_n$  only



needs to satisfy

$$(3.2) \quad \frac{\sum_{n < N} \mathbb{1}_n}{\log N} \rightarrow \infty.$$

But in case of 2-intersectivity and 2-averaging, the structure of  $a_n$  is more complicated: the set  $\mathcal{A}$  contains vectors  $(a_1, a_2, a_3, \dots, a_{N-1})$  given by order 2-correlation

$$a_n = \frac{1}{Q} \sum_{a < Q} g(a) f_1(a+n) f_2(a+2n).$$

The novelty of our paper is that even in this case we manage to show that the  $\epsilon$ -covering number is better than the trivial  $\exp(O(N))$  estimate: we show that for order  $\ell$  correlation sequences, the  $\epsilon$ -covering number is  $\exp(O(N^{1-1/(\ell+1)}))$ , which yields the condition

$$(3.3) \quad \frac{\sum_{n < N} \mathbb{1}_n}{N^{1-1/(\ell+1)}} \rightarrow \infty,$$

appearing in both Theorems A and B. We achieve this by restricting the functions  $f_i$  and  $g$  to thin, *independent* subsets each of a certain appropriate density. The independence of these support sets results in restricting the order  $\ell$  correlation sequences to the *product* of these densities. The fact that we have  $\ell + 1$  factors in an order  $\ell$  correlation sequence explains the appearance of  $\ell + 1$  in the condition eq. (3.3). The origin of the idea of thinning via independent sets is coming from [CG10]. We later noticed that instead of using independent random variables, we can simply use independent residue classes and the Chinese remainder theorem, and this less sophisticated method yields more accurate results.

We do believe, though, that even for order  $\ell > 1$  correlation sequences, the  $\epsilon$ -covering numbers are not greater than some power of  $N$ . Hence the  $\epsilon$ -covering number is always  $\exp(O(\log N))$ , which would give the condition eq. (3.2). This belief is behind Conjectures 2.5 and 2.7.

We close this section by stating that finding optimal  $\epsilon$ -covering numbers is behind many convergence theorems. The general techniques of maximal inequalities, convexity methods (cf. Section 7.3), Fourier transform, spectral theorem, uniform boundedness and transference principles can be viewed as tools to help us reduce the original class of functions for which we need to check convergence.

#### 4. PROOF OF THEOREM A, THE RANDOM INTERSECTIVITY THEOREM

The proof is completely elementary, and, unlike earlier methods, the  $\ell = 2$  case contains all the ingredients of the proof. Hence, we give the detailed proof for  $\ell = 2$ , and only indicate the minor adjustments to be made for  $\ell \geq 3$ .

**4.1. The  $\ell = 2$  case.** As we indicated, we shall work with the weaker assumption in eq. (2.4) which in this case is

$$(4.1) \quad \frac{\sum_{n < N} \mathbb{1}_n}{N^{2/3}} \rightarrow \infty,$$

and we shall prove a little more than required: we shall prove that the set of differences  $r \in R^\omega$  for which  $\overline{d}(A \cap (A-r) \cap (A-2r)) > 0$  is of positive *lower* density in  $R^\omega$ ,

$$(4.2) \quad \liminf_{N \rightarrow \infty} \frac{1}{\sum_{n < N} U_n} \sum_{n < N} U_n \cdot \overline{d}(A \cap (A-n) \cap (A-2n)) > 0.$$

By the strong law of large numbers, Theorem 6.1,

$$\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} U_n \rightarrow 1 \text{ with probability 1,}$$

so eq. (4.2) is equivalent with

$$\liminf_{N \rightarrow \infty} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} U_n \cdot \overline{d}(A \cap (A-n) \cap (A-2n)) > 0.$$

Let us consider the *expectation* of the averages  $\frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} U_n \cdot \bar{d}(A \cap (A-n) \cap (A-2n))$ ,

$$\frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} \mathbb{1}_n \cdot \bar{d}(A \cap (A-n) \cap (A-2n)).$$

Using that the sequence  $(\mathbb{1}_n)$  is decreasing<sup>1</sup>, summation by parts gives the estimate

$$\liminf_{N \rightarrow \infty} \frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} \mathbb{1}_n \cdot \bar{d}(A \cap (A-n) \cap (A-2n)) \geq \liminf_N \frac{1}{N} \sum_{n<N} \bar{d}(A \cap (A-n) \cap (A-2n)).$$

Now, Szemerédi's (actually, in this  $\ell = 2$  case, Roth's) result has been strengthened by [Fur77] to the extent that for any  $A$  (of positive upper density) the *lower* density of the differences of three term arithmetic progressions in  $A$  is positive

$$(4.3) \quad \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n<N} \bar{d}(A \cap (A-n) \cap (A-2n)) > 0.$$

In fact, we need a more precise version of this result: Examining Furstenber's transfer principle in [Fur77], it is clear that for a fixed  $A$  of positive upper density, there is a *single* sequence  $M_1, M_2, \dots, M_k, \dots$  so that for every  $n$  the limit

$$\rho(A \cap (A-n) \cap (A-2n)) = \lim_{k \rightarrow \infty} \frac{|\{a \mid a \in A \cap (A-n) \cap (A-2n), a < M_k\}|}{M_k}$$

exists and satisfies

$$(4.4) \quad \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n<N} \rho(A \cap (A-n) \cap (A-2n)) > 0.$$

Those who seek a more elementary reason for this can use the method of [Var59] to establish eq. (4.4) and hence eq. (4.3).

It is then enough to prove that the difference between the averages along the random sequence and the expectation of the averages goes to 0

$$(4.5) \quad \lim_{N \rightarrow \infty} \sup_A \left| \frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} (U_n - \mathbb{1}_n) \cdot \rho(A \cap (A-n) \cap (A-2n)) \right| = 0 \quad \text{with probability 1,}$$

where in  $\sup_A$  we take the supremum over *all* subsets of the positive integers.

By the assumption in eq. (4.1), we can choose a sequence  $\epsilon_1, \epsilon_2, \dots, \epsilon_N, \dots$  of positive numbers so that

$$(4.6) \quad \epsilon_N \rightarrow 0 \text{ and } \frac{\epsilon_N^2 \sum_{n<N} \mathbb{1}_n}{N^{2/3}} > C, \text{ with some constant } C > 0 \text{ to be chosen later, for all } N > N(C).$$

We will show that, for each  $N > N(C)$ , the  $N$ th average is greater than  $\epsilon_N$  with small probability

$$(4.7) \quad P \left( \sup_A \left| \frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} (U_n - \mathbb{1}_n) \cdot \rho(A \cap (A-n) \cap (A-2n)) \right| > \epsilon_N \right) < \exp \left( -c \epsilon_N^2 \sum_{n<N} \mathbb{1}_n \right).$$

This implies eq. (4.5) since, by the choice of  $\epsilon_N$  in eq. (4.6), we have  $\exp(-c \epsilon_N^2 \sum_{n<N} \mathbb{1}_n) < \exp(-C c N^{2/3})$  which is summable in  $N$ , hence we can apply the Borel-Cantelli lemma.

Now, the difficulty in proving eq. (4.7) is the requirement that we need to have the estimate of the average *simultaneously* for all  $A$ . Indeed, the estimate for a *fixed* set  $A$  instead of  $\sup_A$  follows readily from Bernstein's exponential inequality, [Pet75, page 52],

**Lemma 4.1** (Bernstein's exponential inequality). *Let  $X_1, X_2, \dots, X_N$  be independent, mean zero random variables with  $|X_n| \leq K$ .*

*Then we have*

$$P \left( \left| \sum_{n<N} X_n \right| \geq t \right) \leq 2 \max \left\{ \exp \left( -\frac{t^2/4}{\sum_{n<N} \mathbb{E} X_n^2} \right), \exp(-t/(2K)) \right\} \quad \text{for all } t > 0.$$

<sup>1</sup>This is the only place where we use that the  $(\mathbb{1}_n)$  is decreasing.

**Corollary 4.2.** *Let  $\{a_1, a_2, a_3, \dots, a_n, \dots\}$  be a sequence of real or complex numbers with  $|a_n| \leq 1$ . Then for any  $\epsilon, 0 \leq \epsilon < 1$ , we have*

$$P\left(\left|\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot a_n\right| > \epsilon\right) < \exp\left(-\frac{1}{4}\epsilon^2 \sum_{n < N} \mathbb{1}_n\right).$$

*Proof of Corollary 4.2.* Set  $X_n = (U_n - \mathbb{1}_n) \cdot a_n$ . Then  $|X_n| \leq 1$  and

$$\begin{aligned} \mathbb{E}X_n^2 &\leq \mathbb{E}(U_n - \mathbb{1}_n)^2 \\ &\leq \mathbb{1}_n. \end{aligned}$$

Now use Lemma 4.1 with this  $X_n$  and  $t = \epsilon \cdot \sum_{n < N} \mathbb{1}_n$ . □

Using this corollary with  $a_n = \rho\left(A \cap (A - n) \cap (A - 2n)\right)$  and  $\epsilon = \epsilon_N$ , we readily get

$$P\left(\left|\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \rho\left(A \cap (A - n) \cap (A - 2n)\right)\right| > \epsilon_N\right) < \exp\left(-\frac{1}{4}\epsilon_N^2 \sum_{n < N} \mathbb{1}_n\right).$$

How can we get almost the same estimate when, instead a fixed  $A$ , we take  $\sup_A$ ? The best general estimate we have for the supremum of a class of random variables  $Y_k, k \in \mathcal{K}$ , is what is known as the union estimate

$$(4.8) \quad P\left(\sup_{k \in \mathcal{K}} Y_k > \epsilon\right) \leq |\mathcal{K}| \cdot \sup_{k \in \mathcal{K}} P(Y_k > \epsilon).$$

In view of the above, since in our case we would take  $\mathcal{K}$  to be a set of  $A$ 's, our task is to reduce the number of sets  $A$  we need to consider in  $\sup_A$ . Our first step is to show that we can take *finitely many*  $A$ . In calculating  $\rho\left(A \cap (A - n) \cap (A - 2n)\right)$ , we are taking the limit of  $\frac{1}{M_k} \left|\{a \mid a \in A \cap (A - n) \cap (A - 2n), a < M_k\}\right|$  as  $k \rightarrow \infty$ . Our formulas become simpler if we use the indicator  $\mathbb{1}_A$  of  $A$ . We can then write

$$\frac{1}{M_k} \left|\{a \mid a \in A \cap (A - n) \cap (A - 2n), a < M_k\}\right| = \frac{1}{M_k} \sum_{a < M_k} \mathbb{1}_A(a) \mathbb{1}_A(a + n) \mathbb{1}_A(a + 2n).$$

Let us divide up the interval  $[1, M_k)$  of summation into intervals of length  $Q$ , where we'll see that the best choice for the positive integer  $Q$  will be a constant multiple of  $N^{2/3}$ ,

$$\begin{aligned} &\frac{1}{M_k} \sum_{a < M_k} \mathbb{1}_A(a) \mathbb{1}_A(a + n) \mathbb{1}_A(a + 2n) \\ &= \frac{1}{M_k/Q} \sum_{i < M_k/Q} \frac{1}{Q} \sum_{iQ \leq a < (i+1)Q} \mathbb{1}_A(a) \mathbb{1}_A(a + n) \mathbb{1}_A(a + 2n) + O(Q/M_k). \end{aligned}$$

It follows, since  $M_k \rightarrow \infty$  hence the error term  $O(Q/M_k)$  goes to 0, that

$$\begin{aligned} &\left|\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \rho\left(A \cap (A + n) \cap (A + 2n)\right)\right| \\ &\leq \sup_j \left|\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \frac{1}{Q} \sum_{jQ \leq a < (j+1)Q} \mathbb{1}_A(a) \mathbb{1}_A(a + n) \mathbb{1}_A(a + 2n)\right|. \end{aligned}$$

Now, instead of taking the supremum  $\sup_j$  over varying intervals of length  $Q$ , we can take the *fixed* interval  $[0, Q)$ , and take the supremum over indicators over the intervals  $[0, Q)$ ,  $[0, Q + N)$  and  $[0, Q + 2N)$ . For this, simply change variable  $b = a - jQ$ , and define

$$\begin{aligned} f_0(b) &= \mathbb{1}_A(jQ + b) \mathbb{1}_{[0, Q)}(b), \\ f_1(b) &= \mathbb{1}_A(jQ + b) \mathbb{1}_{[0, Q + N)}(b), \\ f_2(b) &= \mathbb{1}_A(jQ + b) \mathbb{1}_{[0, Q + 2N)}(b). \end{aligned}$$

It follows that

$$\begin{aligned} & \sup_j \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \frac{1}{Q} \sum_{jQ \leq a < (j+1)Q} \mathbb{1}_A(a) \mathbb{1}_A(a+n) \mathbb{1}_A(a+2n) \right| \\ & \leq \sup_{f_i \in \mathcal{F}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \frac{1}{Q} \sum_{b < Q} f_0(b) f_1(b+n) f_2(b+2n) \right|, \end{aligned}$$

where  $\mathcal{F}_i = \mathcal{F}_i(N)$  is the family of indicators supported on the interval  $[0, Q + iN)$  for  $i = 0, 1, 2$ ,

$$\mathcal{F}_i = \left\{ f_i \mid f_i : [0, Q + iN) \rightarrow \{0, 1\} \right\}, i = 0, 1, 2.$$

We've reduced the inequality in eq. (4.7) to

$$(4.9) \quad P \left( \sup_{f_i \in \mathcal{F}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \frac{1}{Q} \sum_{a < Q} f_0(a) f_1(a+n) f_2(a+2n) \right| > \epsilon_N \right) < \exp \left( -c \epsilon_N^2 \sum_{n < N} \mathbb{1}_n \right).$$

The important thing is that in  $\sup_{f_i \in \mathcal{F}_i}$ ,  $i = 0, 1, 2$ , we take the supremum over *finite* sets. The cardinality of  $\mathcal{F}_0$  is  $2^Q$  and at this point it is still at our disposal, so it is of no concern right now except we note that  $Q < N$ . But we do have a problem with the cardinalities of  $\mathcal{F}_i$ ,  $i = 1, 2$ : their cardinalities is at least  $2^N = \exp(cN)$ . Applying Corollary 4.2 with  $a_n = \frac{1}{Q} \sum_{a < Q} f_0(a) f_1(a+n) f_2(a+2n)$ , and using the union estimate eq. (4.8), we only get

$$P \left( \sup_{f_i \in \mathcal{F}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \frac{1}{Q} \sum_{a < Q} f_0(a) f_1(a+n) f_2(a+2n) \right| > \epsilon_N \right) \leq \exp(cN) \cdot \exp \left( -\frac{1}{4} \epsilon_N^2 \sum_{n < N} \mathbb{1}_n \right),$$

which is not good enough since the best estimate we have is  $\epsilon_N^2 \sum_{n < N} \mathbb{1}_n > CN^{2/3}$ . To reduce the cardinality of the  $f_1, f_2$  further, we restrict their support, the  $[0, Q + N)$  and  $[0, Q + 2N)$  intervals, to “independent” arithmetic progressions. We choose 2 moduli  $q_1, q_2$ , and we split the support of  $f_i$ , the interval  $[0, Q + iN)$ , according to the residue classes  $(\text{mod } q_i)$ . This results in splitting the supports of the products  $f_1(a+n) f_2(a+2n)$  into  $q_1 q_2$  residue classes: Let  $\chi_i$  be the indicator of the multiples of  $q_i$

$$\chi_i(x) = \begin{cases} 1 & \text{if } q_i \mid x, \\ 0 & \text{otherwise.} \end{cases}$$

Since  $\chi_i(x-r)$  is the indicator of the  $r \pmod{q_i}$  residue class, we can write

$$\begin{aligned} & \frac{1}{Q} \sum_{a < Q} f_0(a) f_1(a+n) f_2(a+2n) \\ (4.10) \quad & = \frac{1}{Q} \sum_{a < Q} f_0(a) \sum_{r_1 < q_1} \chi_1(a+n-r_1) f_1(a+n) \sum_{r_2 < q_2} \chi_2(a+2n-r_2) f_2(a+2n) \\ & = \frac{1}{Q} \sum_{r_1 < q_1} \sum_{r_2 < q_2} \sum_{a < Q} f_0(a) \chi_1(a+n-r_1) f_1(a+n) \chi_2(a+2n-r_2) f_2(a+2n). \end{aligned}$$

We want the  $q_i$  to be as big as possible, but we don't want the number of terms  $q_1 q_2$  in the sum  $\sum_{r_1 < q_1} \sum_{r_2 < q_2}$  to be greater than  $Q$ , so we require

$$(4.11) \quad q_1 q_2 \leq Q.$$

We also need to make sure  $\sum_{a < Q} f_0(a) \chi_1(a+n-r_1) f_1(a+n) \chi_2(a+2n-r_2) f_2(a+2n)$  is bounded independently of  $n$  for fixed  $r_i$ . We can achieve this if the equation

$$(4.12) \quad \chi_1(a+n-r_1) \chi_2(a+2n-r_2) = 1$$

has, say, at most 1 solution in  $a \in [0, Q)$ . We can rewrite eq. (4.12) as simultaneous congruences

$$\begin{aligned} a &\equiv r_1 - n \pmod{q_1}, \\ a &\equiv r_2 - 2n \pmod{q_2}. \end{aligned}$$

Now, if the  $q_i$  are coprimes, then the above system of congruences has a single solution  $\pmod{q_1 q_2}$  by the Chinese remainder theorem. It follows that eq. (4.12) has at most 1 solution in  $a \in [0, Q)$  if

$$Q \leq q_1 q_2.$$

Comparing this with eq. (4.11), we commit to

$$Q = q_1 q_2.$$

We then have

$$\left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \frac{1}{Q} \sum_{a < Q} f_0(a) f_1(a+n) f_2(a+2n) \right|$$

by eq. (4.10) and since  $Q = q_1 q_2$

$$\begin{aligned} &= \left| \frac{1}{Q} \sum_{r_1 < q_1} \sum_{r_2 < q_2} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \sum_{a < Q} f_0(a) \chi_1(a+n-r_1) f_1(a+n) \chi_2(a+2n-r_2) f_2(a+2n) \right| \\ &\leq \sup_{r_i < q_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \sum_{a < Q} f_0(a) \chi_1(a+n-r_1) f_1(a+n) \chi_2(a+2n-r_2) f_2(a+2n) \right|. \end{aligned}$$

It follows, using the union estimate eq. (4.8),

$$\begin{aligned} &P\left(\sup_{f_i \in \mathcal{F}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \sum_{a < Q} f_0(a) f_1(a+n) f_2(a+2n) \right| > \epsilon_N \right) \\ &\leq q_1 q_2 \sup_{r_i < q_i} P\left(\sup_{f_i \in \mathcal{F}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \sum_{a < Q} f_0(a) \chi_1(a+n-r_1) f_1(a+n) \chi_2(a+2n-r_2) f_2(a+2n) \right| > \epsilon_N \right). \end{aligned}$$

The contribution of  $q_1 q_2 = Q < N = \exp(\log N)$  is insignificant to the order of  $\exp(-\epsilon_N^2 \sum_{n < N} \mathbb{1}_n) = \exp(-O(N^{2/3}))$  which we aim to achieve in eq. (4.9), so we ignore it in further considerations. For given  $r_i$ ,  $i = 1, 2$ , the function  $g_i$  defined by  $g_i(x) = \chi(x - r_i) f_i(x)$  is supported on the part of the residue class  $r_i \pmod{q_i}$  that falls into the interval  $[0, Q + iN)$ . Denote by  $\mathcal{G}_i$  those functions from  $\mathcal{F}_i$  which are supported on the residue class  $r_i \pmod{q_i}$

$$\mathcal{G}_i = \left\{ g_i \mid g_i : [0, Q + iN) \rightarrow \{0, 1\}, g_i(x) \neq 0 \implies x \equiv r_i \pmod{q_i} \right\}.$$

Since the cardinality of the residue class  $r_i \pmod{q_i}$  that falls into the interval  $[0, Q + iN)$  is not more than  $(Q + iN)/q_i < 3N/q_i$ , we have

$$(4.13) \quad |\mathcal{G}_i| \leq 2^{2N/q_i}, i = 1, 2.$$

Our remaining task is to estimate, using the convenient notation  $\mathcal{G}_0 = \mathcal{F}_0$ ,

$$P\left(\sup_{g_i \in \mathcal{G}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \sum_{a < Q} g_0(a) g_1(a+n) g_2(a+2n) \right| > \epsilon_N \right).$$

In the above,  $|\sum_{a < Q} g_0(a) g_1(a+n) g_2(a+2n)|$  is bounded by 1 for each  $n$ . This is because  $q_1 q_2 = Q$ ,  $q_1$  and  $q_2$  are coprimes, hence, by the Chinese remainder theorem, there is one  $a < Q$  for which  $a+n \equiv r_1 \pmod{q_1}$  and  $a+2n \equiv r_2 \pmod{q_2}$ . Clearly

$$|\mathcal{G}_0| = 2^Q,$$

hence, using eq. (4.13),

$$|\mathcal{G}_0 \times \mathcal{G}_1 \times \mathcal{G}_2| \leq 2^{Q+2N/q_1+2N/q_2}.$$

We can now estimate, using the union estimate eq. (4.8) and Corollary 4.2 with  $a_n = \sum_{a < Q} g_0(a) g_1(a+n) g_2(a+2n)$ , as

$$\begin{aligned} &P\left(\sup_{g_i \in \mathcal{G}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \sum_{a < Q} g_0(a) g_1(a+n) g_2(a+2n) \right| > \epsilon_N \right) \\ (4.14) \quad &\leq 2^{Q+2N/q_1+2N/q_2} \cdot \exp\left(-\frac{1}{4} \epsilon_N^2 \sum_{n < N} \mathbb{1}_n\right) \\ &\leq \exp\left((Q+2N/q_1+2N/q_2) - \frac{1}{4} \epsilon_N^2 \sum_{n < N} \mathbb{1}_n\right), \end{aligned}$$

where we used that  $\log 2 < 1$ . In order to have a nontrivial estimate, we need to have something like

$$(4.15) \quad Q + \frac{2N}{q_1} + \frac{2N}{q_2} \leq \frac{1}{5} \cdot \epsilon_N^2 \sum_{n < N} \mathbb{1}_n.$$

Treating the left hand side as if  $q_i$  and  $Q$  were real variables, since  $Q = q_1 q_2$ , the left hand side has a minimum at  $Q = cN^{2/3}$ , and  $q_1 = q_2 = cN^{1/3}$ . We certainly can choose coprime integers  $q_i$  which nearly satisfy these requirements. With these choices eq. (4.15) becomes

$$cN^{2/3} \leq \frac{1}{5} \cdot \epsilon_N^2 \sum_{n < N} \mathbb{1}_n.$$

We certainly can achieve this by taking  $C$  large enough in eq. (4.6). With these choices of  $C, Q, q_i$ , the estimate in eq. (4.14) becomes

$$P\left(\sup_{g_i \in \mathcal{G}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \sum_{a < Q} g_0(a) g_1(a+n) g_2(a+2n) \right| > \epsilon_N \right) \leq \exp(-cN^{2/3}),$$

finishing our proof.

**4.2. The general  $\ell$  case.** There are only three remarks we need to make here, and then the proof goes through verbatim.

The first remark is that in [Fur77] eq. (3.2) is proved for all  $\ell$ :

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} \bar{d}\left(A \cap (A-n) \cap \dots \cap (A-(\ell-1)n) \cap (A-\ell n)\right) > 0.$$

For an elementary proof: N. Hegyvári pointed out to us that the method of [Var59] can be used for the same conclusion<sup>2</sup>. The slightly strengthened version we really need, and which we described in eq. (4.4), is also valid.

Second, this time we need to choose the  $(\epsilon_N)$  so that

$$\epsilon_N \rightarrow 0 \text{ and } \frac{\epsilon_N^2 \sum_{n < N} \mathbb{1}_n}{N^{1-1/(\ell+1)}} > C, \text{ with some constant } C > 0 \text{ to be chosen later, for all } N > N(C).$$

For the third remark, let us look at the essential inequality we really need to prove: with a  $Q < N$  to be chosen later, we need to prove the following analog of eq. (4.9)

$$P\left(\sup_{f_i \in \mathcal{F}_i} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot \frac{1}{Q} \sum_{a < Q} f_0(a) f_1(a+n) \dots f_{\ell-1}(a+(\ell-1)n) f_\ell(a+\ell n) \right| > \epsilon_N \right) < \exp\left(-c\epsilon_N^2 \sum_{n < N} \mathbb{1}_n\right),$$

where

$$\mathcal{F}_i = \{f_i \mid f_i : [0, Q+iN) \rightarrow \{0, 1\}\}, i = 0, 1, 2, \dots, \ell.$$

This time, we need to slim down the supports of  $\ell$  functions  $f_1, f_2, \dots, f_\ell$ . Correspondingly, we choose  $\ell$  pairwise coprime moduli  $q_1, q_2, \dots, q_\ell$  to split the support of  $f_i$ , the interval  $[0, Q+iN)$ , according to the residue classes  $(\text{mod } q_i)$ . Proceeding as in the case  $\ell = 2$ , we get the requirements

$$(4.16) \quad q_1 q_2 \dots q_\ell = Q,$$

and

$$(4.17) \quad \frac{\ell N}{q_1} + \frac{\ell N}{q_2} + \dots + \frac{\ell N}{q_\ell} + Q \leq \frac{1}{5} \epsilon_N^2 \sum_{n < N} \mathbb{1}_n.$$

If the  $q_i$  and  $Q$  were real variables, the left hand side would be minimized when

$$Q = cN^{\ell/(\ell+1)},$$

and the  $q_i$  were equal to each other. By eq. (4.16) this means

$$(4.18) \quad q_i = Q^{1/\ell} = cN^{1/(\ell+1)}.$$

<sup>2</sup>In [Var59], only the  $\ell = 2$  case is treated, but the method works for any  $\ell$ . For a proof for any  $\ell$ , see Theorem 8' in <http://terrytao.wordpress.com/2008/08/30/the-correspondence-principle-and-finitary-ergodic-theory/>.

With these values, eq. (4.17) becomes, with some  $c$  depending on  $\ell$  only,

$$(4.19) \quad cN^{1-1/(\ell+1)} < \epsilon_N^2 \sum_{n < N} \mathbb{1}_n,$$

which suggested the assumption in eq. (2.4). The only question remains if we can choose pairwise coprime  $q_i$ 's near the optimal value  $N^{1/(\ell+1)}$  of eq. (4.18)? To do this, choose each  $q_i$  to be a prime number with  $N^{1/(\ell+1)} < q_i < 2N^{1/(\ell+1)}$ . This is possible for large enough  $N$  by the prime number theorem. The much simpler Chebyshev estimate  $\frac{1}{2}t/\log t < \pi(t) < 2t/\log t$  is also enough, but then the primes would satisfy  $N^{1/(\ell+1)} < q_i < 5N^{1/(\ell+1)}$ .

## 5. PROOF OF THEOREM B, THE RANDOM AVERAGING THEOREM

Similarly to the proof of Theorem A, we just explain the idea of the proof for  $\ell = 2$ . Since the proof is very similar to the proof of Theorem A, we just point out the differences between the arguments.

**5.1. The  $\ell = 2$  case.** We want to show that with probability 1 in any dynamical system

$$(5.1) \quad \frac{1}{\sum_{n < N} U_n} \sum_{n < N} U_n \cdot T^n F_1 \cdot T^{2n} F_2 \text{ converge in } L^1 \text{ norm.}$$

Approximating  $F_i$  in  $L^1$  norm by linear combinations of indicator functions, we see it's enough to show convergence for indicators  $F_i$  of measurable subsets of  $X$ . By the strong law of large numbers, Theorem A.1, we have

$$\frac{\sum_{n < N} U_n}{\sum_{n < N} \mathbb{1}_n} \rightarrow 1 \text{ with probability 1,}$$

hence eq. (5.1) is equivalent to

$$\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} U_n \cdot T^n F_1 \cdot T^{2n} F_2 \text{ converge in } L^1 \text{ norm.}$$

The expectation of the averages  $\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} U_n \cdot T^n F_1 \cdot T^{2n} F_2$  is

$$(5.2) \quad \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} \mathbb{1}_n \cdot T^n F_1 \cdot T^{2n} F_2.$$

Summation by parts shows that the averages in eq. (5.2) converge to the same limit as the double ergodic averages  $\frac{1}{N} \sum_{n < N} T^n F_1 \cdot T^{2n} F_2$ . That these double averages converge in norm was proved by [Fur77], so it is enough to prove that the differences between the random averages and their expectations go to 0 in  $L^1$  norm

$$\lim_{N \rightarrow \infty} \sup_{(X, T, \mu), F_i} \int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^n F_1 \cdot T^{2n} F_2 \right| = 0 \text{ with probability 1,}$$

where in  $\sup_{(X, T, \mu), F_i}$  we take the supremum over all dynamical systems  $(X, T, \mu)$  and indicators  $F_i$ ,  $i = 1, 2$ .

We first reduce all to a single dynamical system using an appropriate version of the transference principle of A. Calderón. Since  $T$  is measure preserving, we have, for every integer  $a$

$$\int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^n F_1 \cdot T^{2n} F_2 \right| = \int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^{a+n} F_1 T^{a+2n} F_2 \right|.$$

Average this in  $a$  over the interval  $[0, Q]$ , where  $Q$  is the same as in the proof of Theorem A in the  $\ell = 2$  case, so its best choice will be  $Q \sim N^{2/3}$ . We get

$$\int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^n F_1 \cdot T^{2n} F_2 \right| = \int_X \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^{a+n} F_1 T^{a+2n} F_2 \right|.$$

It follows that

$$(5.3) \quad \sup_{F_i} \int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^n F_1 \cdot T^{2n} F_2 \right| \leq \int_X \sup_{F_i} \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^{a+n} F_1 T^{a+2n} F_2 \right|.$$

Define the  $\mathbb{N} \rightarrow \{0, 1\}$  indicators  $f_1, f_2$  by

$$\begin{aligned} f_1(a) &= f_1(x, a) = T^a F_1(x) \mathbb{1}_{[0, Q+N)}(a), \\ f_2(a) &= f_2(x, a) = T^a F_2(x) \mathbb{1}_{[0, Q+2N)}(a). \end{aligned}$$

Then for the values of  $a$  and  $n$  considered above we have  $T^{a+n} F_1(x) = f_1(a+n)$  and  $T^{a+2n} F_1(x) = f_2(a+2n)$ . Denote by  $\mathcal{F}_i$  the family of indicators supported on the interval  $[0, Q+iN)$

$$\mathcal{F}_i = \{f_i \mid f_i : [0, Q+iN) \rightarrow \{0, 1\}\}.$$

We then have the pointwise estimate

$$\begin{aligned} & \sup_{F_i} \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^{a+n} F_1(x) T^{a+2n} F_2(x) \right| \\ & \leq \sup_{f_i \in \mathcal{F}_i} \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot f_1(a+n) f_2(a+2n) \right|. \end{aligned}$$

Integrating this over the probability space  $X$ , and using eq. (5.3), we get

$$\begin{aligned} (5.4) \quad & \sup_{(X, T, \mu)} \sup_{F_i} \int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot T^n F_1 T^{2n} F_2 \right| \\ & \leq \int_X \sup_{f_i \in \mathcal{F}_i} \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot f_1(a+n) f_2(a+2n) \right| \\ & = \sup_{f_i \in \mathcal{F}_i} \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot f_1(a+n) f_2(a+2n) \right| \end{aligned}$$

The important thing to notice is that the final estimate of eq. (5.4) doesn't depend on the dynamical system  $(X, T, \mu)$  anymore and, furthermore, in  $\sup_{f_i \in \mathcal{F}_i}$  we take the supremum over finitely many functions  $f_i$ ,  $i = 1, 2$ . Our remaining task is to show that

$$(5.5) \quad \lim_{N \rightarrow \infty} \sup_{f_i \in \mathcal{F}_i} \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot f_1(a+n) f_2(a+2n) \right| = 0 \text{ with probability 1.}$$

Let us write the above average in dual form: Let  $\mathcal{G} = \mathcal{G}(N)$  be the family of  $\pm 1$ -valued functions supported on the interval  $[0, Q)$

$$\mathcal{G} = \{g \mid g : [0, Q) \rightarrow \{-1, 1\}\}.$$

We then have

$$\begin{aligned} (5.6) \quad & \sup_{f_i \in \mathcal{F}_i} \frac{1}{Q} \sum_{a < Q} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot f_1(a+n) f_2(a+2n) \right| \\ & = \sup_{f_i \in \mathcal{F}_i, g \in \mathcal{G}} \frac{1}{Q} \sum_{a < Q} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \cdot g(a) f_1(a+n) f_2(a+2n) \\ & = \sup_{f_i \in \mathcal{F}_i, g \in \mathcal{G}} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \frac{1}{Q} \sum_{a < Q} g(a) f_1(a+n) f_2(a+2n). \end{aligned}$$

The only difference between the right hand side of eq. (5.6) above and

$$(5.7) \quad \sup_{f_i \in \mathcal{F}_i} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \frac{1}{Q} \sum_{a < Q} f_0(a) f_1(a+n) f_2(a+2n)$$

of eq. (4.9) is that  $\mathcal{G}$  contains  $\pm 1$ -valued functions while the  $\mathcal{F}_0$  of eq. (5.7) contains indicators. But this difference doesn't change anything in the proof, since the only thing we use about the functions in  $\mathcal{G}$  is that they take on two values. So the method used to establish eq. (4.9) works equally well to prove, for all large enough  $N$ ,

$$(5.8) \quad P \left( \sup_{f_i \in \mathcal{F}_i, g \in \mathcal{G}} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \frac{1}{Q} \sum_{a < Q} g(a) f_1(a+n) f_2(a+2n) > \epsilon_N \right) < \exp \left( -c \epsilon_N^2 \sum_{n < N} \mathbb{1}_n \right),$$



where  $\epsilon_N$  is chosen to satisfy eq. (4.6). But this choice of  $\epsilon_N$  implies that the right hand side of eq. (5.8) is summable and hence, by the Borel-Cantelli lemma, we have eq. (5.5).

**5.2. The general  $\ell$  case.** Our remarks are similar to the three remarks we made in Section 4.2. The only difference is in the first remark: While in the  $\ell = 2$  case, we invoked Furstenberg's double convergence theorem, this time we need to refer to the "multiple mean ergodic theorem" of [HK05b] (cf [Zie07] for another proof). The multiple mean ergodic theorem says that the set of positive integers is  $\ell$ -averaging, that is the averages

$$\frac{1}{N} \sum_{n < N} T^n F_1 \cdot T^{2n} F_2 \cdots T^{\ell n} F_\ell$$

converge in  $L^1$  norm for bounded  $F_i$ . Then the proof proceeds to show that the averages along the random sequence converge to the same limit as the multiple averages above.

**5.3. Pointwise convergence.** The main philosophy on pointwise convergence for *any* averages (not just those in our paper) is that if mean convergence takes place with an appropriate speed, then we also have pointwise convergence. It is enough to prescribe the speed for a *subsequence* of the averages

$$(5.9) \quad \frac{1}{\sum_{n < N} U_n} \sum_{n < N} (U_n - \mathbb{U}_n) T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x).$$

What is the sparsest subsequence of the averages the convergence of which implies the convergence of all the averages? Actually, we are going to consider a *class* of subsequences: for  $\sigma > 1$ , let  $\{N_1 < N_2 < N_3 < \dots\}$  be such that

$$\sum_{n < N_i} \mathbb{U}_n \approx \sigma^i.$$

Roughly speaking, for each  $\sigma > 1$ , the  $N_i$  defined above is a sequence along which the averages could still change significantly, but as  $\sigma$  gets closer and closer to 1, this change is as small as we want. To understand why, consider the simplest random average relevant to our problem

$$\frac{1}{\sum_{n < N} \mathbb{U}_n} \sum_{n < N} U_n.$$

The  $N_i$ th sum  $\sum_{n < N_i} U_n$  has an expected mass of  $\sum_{n < N_i} \mathbb{U}_n \approx \sigma^i$  and the  $N_{i+1}$ st sum  $\sum_{n < N_{i+1}} U_n$  has an expected mass of  $\sum_{n < N_{i+1}} \mathbb{U}_n \approx \sigma^{i+1}$ . Hence the difference  $\sum_{N_i \leq n < N_{i+1}} U_n$  has an expected mass of  $\sum_{N_i \leq n < N_{i+1}} \mathbb{U}_n \approx \sigma^{i+1} - \sigma^i$ . What is the proportion of this increase in mass to  $\sigma^i$ , the expected mass of the  $N_i$ th average? It is  $(\sigma^{i+1} - \sigma^i) / \sigma^i = \sigma - 1$ . This means, the proportion of this difference is a fixed positive number (independent of  $i$ ), so if changes happen in this difference, it may affect averages that are between the  $N_i$ th and  $N_{i+1}$ st average. On the other hand, the proportion  $\sigma - 1$  of this difference goes to 0 as  $\sigma$  gets closer to 1. We then expect that if the  $N_i$ th average and the  $N_{i+1}$ st averages are both close to 1, and if  $N$  is between  $N_i$  and  $N_{i+1}$ , then the  $N$ th average will also be close to 1, and the difference will not be much more than  $\sigma - 1$ .

Let us now carry out this argument for our averages precisely. Our random average in eq. (5.9) is of the form

$$\frac{1}{\sum_{n < N} U_n} \sum_{n < N} (U_n - \mathbb{U}_n) G_n$$

for some uniformly bounded sequence of functions  $G_1, G_2, G_3, \dots$ . We can assume  $\|G_n\|_\infty \leq 1$ . By the strong law of large numbers, Theorem A.1, with probability 1, we can divide by the expectation  $\sum_{n < N} \mathbb{U}_n$  instead of  $\sum_{n < N} U_n$

$$\frac{1}{\sum_{n < N} \mathbb{U}_n} \sum_{n < N} (U_n - \mathbb{U}_n) G_n.$$

For  $\sigma > 1$ , we choose the subsequence  $I_\sigma = \{N_1 < N_2 < N_3 < \dots\}$  so that  $\sum_{n < N_i} \mathbb{U}_n \approx \sigma^i$ ,

$$\sum_{n < N_i} \mathbb{U}_n \geq \sigma^i \text{ but } \sum_{n < N} \mathbb{U}_n < \sigma^i \text{ for } N < N_i.$$

In words:  $N_i$  is the smallest index  $N$  for which  $\sum_{n < N} \mathbb{U}_n \geq \sigma^i$ . That  $I_\sigma$  is well defined and infinite follows from the dissipative assumption  $\sum_n \mathbb{U}_n = \infty$ . Clearly

$$(5.10) \quad \sum_{n < N_i} \mathbb{U}_n = \sigma^i + O(1) \text{ for all } i = 1, 2, 3, \dots$$

Note that we also have that  $\sum_{n < N_i} U_n \approx \sigma^i$ . More precisely, we have, by the strong law of large numbers, Theorem A.1, and by eq. (5.10),

$$\sum_{n < N_i} U_n = \sigma^i + o(\sigma^i) \text{ with probability 1.}$$

Let us see how convergence along the subsequences  $I_\sigma$ 's,  $\sigma > 1$ , imply convergence along the full sequence  $N = 1, 2, 3, \dots$ . Assume that for each fixed  $\sigma > 1$ , there is  $\Omega_\sigma$  with  $P(\Omega_\sigma) = 1$  so that

$$(5.11) \quad \lim_{i \rightarrow \infty} \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \sum_{n < N_i} (U_n(\omega) - \mathbb{1}_n) G_n(x) = 0 \text{ for } \omega \in \Omega_\sigma \text{ and for a.e. } x \in X.$$

We want to show that then for each fixed  $\sigma > 1$ , there is  $\Omega'_\sigma$  with  $P(\Omega'_\sigma) = 1$  so that

$$(5.12) \quad \limsup_{N \rightarrow \infty} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n(\omega) - \mathbb{1}_n) G_n(x) \right| \leq 4(\sigma - 1) \text{ for } \omega \in \Omega'_\sigma \text{ and for a.e. } x \in X.$$

It's important to emphasize that the construction of the set  $\Omega'_\sigma$  is independent of  $x \in X$ , and that the only property of the sequence  $G_1(x), G_2(x), \dots, G_n(x), \dots$  we use is that it's bounded by 1. For this reason, we suppress the argument  $x$  in  $G_n(x)$ .

Now, if we choose a *sequence* of  $\sigma$ 's which go to 1, then the intersection  $\Omega'$  of the corresponding sets  $\Omega'_\sigma$  for which eq. (5.12) holds will give convergence to 0,

$$\lim_{N \rightarrow \infty} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n(\omega) - \mathbb{1}_n) G_n = 0 \text{ for } \omega \in \Omega'.$$

So fix  $\sigma > 1$ . By the assumption in eq. (5.11), eq. (5.12) follows if we show that there is  $\Omega'_\sigma$  with  $P(\Omega'_\sigma) = 1$  so that

$$(5.13) \quad \limsup_{i \rightarrow \infty} \max_{N_i \leq N < N_{i+1}} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n(\omega) - \mathbb{1}_n) G_n - \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \sum_{n < N_i} (U_n(\omega) - \mathbb{1}_n) G_n \right| \leq 4(\sigma - 1) \text{ for } \omega \in \Omega'_\sigma.$$

For a given  $N$ , choose the  $i$  so that

$$N_i \leq N < N_{i+1}.$$

To estimate the difference between the  $N$ th and  $N_i$ th average, denote

$$H_n = (U_n(\omega) - \mathbb{1}_n) G_n,$$

and write

$$(5.14) \quad \begin{aligned} & \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} H_n - \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \sum_{n < N_i} H_n \\ &= \left( \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} H_n - \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N_i} H_n \right) + \left( \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N_i} H_n - \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \sum_{n < N_i} H_n \right) \\ &= \left( \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{N_i \leq n < N} H_n \right) + \left( \sum_{n < N_i} H_n \right) \left( \frac{1}{\sum_{n < N} \mathbb{1}_n} - \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \right). \end{aligned}$$

To estimate the two terms in eq. (5.14), we need the following consequence of the strong law of large numbers, Theorem A.1,

$$(5.15) \quad \lim_{i \rightarrow \infty} \frac{\sum_{N_i \leq n < N_{i+1}} U_n(\omega)}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} = 1 \text{ with probability 1.}$$

To see eq. (5.15), write as

$$\frac{\sum_{N_i \leq n < N_{i+1}} U_n(\omega)}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} = \frac{1}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \left( \sum_{n < N_{i+1}} U_n(\omega) - \sum_{n < N_i} U_n(\omega) \right).$$

Writing

$$\frac{1}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \sum_{n < N_{i+1}} U_n(\omega) = \frac{\sum_{n < N_{i+1}} \mathbb{1}_n}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \cdot \frac{1}{\sum_{n < N_{i+1}} \mathbb{1}_n} \sum_{n < N_{i+1}} U_n(\omega),$$

we see that by the strong law of large numbers, we have, with probability 1,

$$\lim_{i \rightarrow \infty} \frac{1}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \sum_{n < N_{i+1}} U_n(\omega) = \lim_{i \rightarrow \infty} \frac{\sum_{n < N_{i+1}} \mathbb{1}_n}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n}$$

using eq. (5.10)

$$\begin{aligned} &= \lim_{i \rightarrow \infty} \frac{\sigma^{i+1} + O(1)}{\sigma^{i+1} - \sigma^i + O(1)} \\ &= \frac{\sigma}{\sigma - 1}. \end{aligned} \quad (5.16)$$

Similarly

$$\begin{aligned} \lim_{i \rightarrow \infty} \frac{1}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \sum_{n < N_i} U_n(\omega) &= \lim_{i \rightarrow \infty} \frac{\sum_{n < N_i} \mathbb{1}_n}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \\ &= \frac{1}{\sigma - 1}. \end{aligned} \quad (5.17)$$

Now take the difference of eqs. (5.16) and (5.17) to get our claim in eq. (5.15).

Let's go back to our main task of estimating the two terms in eq. (5.14). Estimate as

$$\begin{aligned} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{N_i \leq n < N} H_n \right| &\leq \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \sum_{N_i \leq n < N_{i+1}} |H_n| \\ &\leq \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \sum_{N_i \leq n < N_{i+1}} (U_n(\omega) + \mathbb{1}_n) \\ &= \frac{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n}{\sum_{n < N_i} \mathbb{1}_n} \cdot \frac{1}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \sum_{N_i \leq n < N_{i+1}} (U_n(\omega) + \mathbb{1}_n). \end{aligned}$$

Take “ $\limsup_{i \rightarrow \infty} \max_{N_i \leq n < N_{i+1}}$ ” of both ends, to get

$$\begin{aligned} \limsup_{i \rightarrow \infty} \max_{N_i \leq n < N_{i+1}} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{N_i \leq n < N} H_n \right| \\ \leq \limsup_{i \rightarrow \infty} \frac{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n}{\sum_{n < N_i} \mathbb{1}_n} \cdot \frac{1}{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n} \sum_{N_i \leq n < N_{i+1}} (U_n(\omega) + \mathbb{1}_n) \end{aligned}$$

using eq. (5.15)

$$= 2 \cdot \lim_{i \rightarrow \infty} \frac{\sum_{N_i \leq n < N_{i+1}} \mathbb{1}_n}{\sum_{n < N_i} \mathbb{1}_n}$$

using eq. (5.10)

$$\begin{aligned} &= 2 \cdot \lim_{i \rightarrow \infty} \frac{\sigma^{i+1} - \sigma^i + O(1)}{\sigma^i + O(1)} \\ &= 2(\sigma - 1). \end{aligned} \quad (5.18)$$

Similarly

$$\limsup_{i \rightarrow \infty} \max_{N_i \leq n < N_{i+1}} \left| \left( \sum_{n < N_i} H_n \right) \left( \frac{1}{\sum_{n < N} \mathbb{1}_n} - \frac{1}{\sum_{n < N_i} \mathbb{1}_n} \right) \right| \leq 2(\sigma - 1). \quad (5.19)$$

The estimates in eqs. (5.18) and (5.19) imply eq. (5.13).

The consequence of this lacunary subsequence trick is that for the averages in eq. (5.9), it is enough to prove, for all  $\sigma > 1$ , almost sure convergence along the subsequence  $I_\sigma$ . This is accomplished if we ensure that the sequence  $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n, \dots\}$  satisfies

$$\sum_{N \in I_\sigma} \epsilon_N < \infty \text{ for all } \sigma > 1 \quad (5.20)$$

in addition to the requirement in eq. (4.19). To see this implication of the condition in eq. (5.20), note first that by the Borel-Cantelli lemma we have, for almost every  $\omega$ ,

$$\sup_{(X,T,\mu),F_i} \int_X \left| \frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} (U_n(\omega) - \mathbb{1}_n) \cdot T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x) \right| \leq \epsilon_N \text{ for } N \geq N(\omega).$$

By eq. (5.20), we have, for almost every  $\omega$ ,

$$\begin{aligned} & \sup_{(X,T,\mu),F_i} \int_X \sum_{N \in I_\sigma} \left| \frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} (U_n(\omega) - \mathbb{1}_n) \cdot T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x) \right| \\ & \leq \sum_{N \in I_\sigma} \sup_{(X,T,\mu),F_i} \int_X \left| \frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} (U_n(\omega) - \mathbb{1}_n) \cdot T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x) \right| \\ & < \infty. \end{aligned}$$

This implies that the series  $\sum_{N \in I_\sigma} \left| \frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} (U_n(\omega) - \mathbb{1}_n) \cdot T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x) \right|$  is convergent for almost every  $x$ , and hence  $\frac{1}{\sum_{n<N} \mathbb{1}_n} \sum_{n<N} (U_n(\omega) - \mathbb{1}_n) \cdot T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x) \rightarrow 0$  for almost every  $x$  as  $N \in I_\sigma$  goes to  $\infty$ .

To ensure eq. (5.20), note that in our case, since we at least assume  $\frac{\sum_{n<N} \mathbb{1}_n}{N^{1-1/(\ell+1)}} \rightarrow \infty$ , the subsequence  $I_\sigma$  does not increase faster to  $\infty$  than a geometric progression. So a near optimal choice for  $\epsilon_N$  to ensure eq. (5.20) is  $\epsilon_N = \log^{-(1+\delta)} N$  for some fixed positive  $\delta$ .

Putting all this together, we get the following theorem for the case when the  $F_i$  are indicator functions

**Theorem D** (Random pointwise  $\ell$ -averaging theorem). *Let  $\ell$  be a positive integer and suppose that for some positive  $\delta$  we have*

$$(5.21) \quad \mathbb{1}_n \cdot \frac{n^{1/(\ell+1)}}{\log^{2+\delta} n} \rightarrow \infty.$$

*Then, with probability 1, in any dynamical system  $(X, T)$ ,*

$$(5.22) \quad \lim_{N \rightarrow \infty} \frac{1}{\sum_{n<N} U_n(\omega)} \sum_{n<N} (U_n(\omega) - \mathbb{1}_n) \cdot T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x) = 0 \text{ for a.e. } x \in X,$$

*for bounded functions  $F_i$ .*

The main utility of this theorem is that once we have almost everywhere convergence for the multiple averages

$$(5.23) \quad \frac{1}{N} \sum_{n<N} T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x),$$

then the averages

$$\frac{1}{\sum_{n<N} U_n} \sum_{n<N} U_n \cdot T^n F_1(x) \cdot T^{2n} F_2(x) \cdots T^{\ell n} F_\ell(x)$$

along the random sequence also converge and to the same limit for almost every  $x$ . So far, the a.e. convergence of eq. (5.23) in *every* dynamical system is known only in case  $\ell = 1, 2$ . For  $\ell = 1$ , this is the pointwise ergodic theorem, for  $\ell = 2$  it was proved by [Bou90]. Moreover, the pointwise convergence of the averages in eq. (5.23) for any  $\ell$  is known in a great variety of particular dynamical systems such as in

- discrete spectrum or quasi-discrete spectrum dynamical systems;
- K-systems (cf. [DL96]);
- a large class of skew-product extensions of ergodic compact group translations (cf. [Les93]);
- a lot of dynamical systems with a compact topological structure where *everywhere* convergence can be proved for continuous functions, hence the general pointwise ergodic theorem follows by a maximal inequality. An example for this that we haven't mentioned is a horocycle flow on a compact surface with negative curvature, or, more generally, unipotent transformations (Ratner's theory).

So far, we showed how to prove Theorem D for indicator functions. The extension to bounded functions is more or less standard: Since we have a.e. convergence for a dense set of functions, namely for linear combinations of indicators, we just need a "maximal inequality" argument. As we will see, in this maximal

inequality argument, we will work with a set of  $\omega$  which satisfy conditions that are *independent* of the dynamical system. The two main conditions are the a.e. convergence of the averages in eq. (5.22) when the  $F_i$  are indicators, and the maximal inequality in eq. (5.25) below. We will also use the strong law of large numbers, Theorem A.1.

The way we'll do the extension of a.e. convergence to bounded functions is that we first extend a.e. convergence to all bounded  $F_1$ , then we extend it to all bounded  $F_1, F_2$ , then to all bounded  $F_1, F_2, F_3$ , etc.

For this argument, it is convenient to introduce three notations: one for our averages, normalized by  $\sum_{n < N} \mathfrak{u}_n$  instead of  $\sum_{n < N} U_n$

$$A_N(F_1, F_2, \dots, F_\ell) = \frac{1}{\sum_{n < N} \mathfrak{u}_n} \sum_{n < N} (U_n - \mathfrak{u}_n) \cdot T^n F_1 \cdot T^{2n} F_2 \cdots T^{\ell n} F_\ell,$$

and two for the 1-linear, dominating averages

$$B_N F = \frac{1}{\sum_{n < N} \mathfrak{u}_n} \sum_{n < N} \mathfrak{u}_n \cdot |T^n F|, \quad C_N F = \frac{1}{\sum_{n < N} \mathfrak{u}_n} \sum_{n < N} U_n \cdot |T^n F|.$$

A summation by parts argument shows that  $B_N$  is pointwise dominated by the usual ergodic averages  $\frac{1}{N} \sum_{n < N} |T^n F|$ , and hence we have the maximal inequality for the  $B_N$

$$(5.24) \quad \int_X \left( \sup_N B_N F \right)^2 \leq c \int_X F^2.$$

With probability 1, we also have a maximal inequality for the  $C_N$

$$(5.25) \quad \int_X \left( \sup_N C_N F \right)^2 \leq c_\omega \int_X F^2.$$

This was established by [Bou88] for a much wider range of  $\mathfrak{u}_n$  than we consider in this theorem: they just need to satisfy

$$\frac{\sum_{n < N} \mathfrak{u}_n}{(\log \log N)^{1+\delta} \log N} \rightarrow \infty \text{ for some positive } \delta.$$

That this extra  $\log \log N$  factor is *necessary* for the maximal inequality and a.e. convergence, compared to the lone  $\log N$  needed for mean convergence, was proved in [JLW99].

Let now  $F_1$  be a function bounded by 1, and assume  $F_2, F_3, \dots, F_\ell$  are indicators. We want to show that, with probability 1, the oscillation of the  $A_N$  is 0 almost surely

$$\mu \left( \limsup_N |A_N(F_1, F_2, \dots, F_\ell)| = 0 \right) = 1.$$

We show this by showing that for any  $\epsilon > 0$

$$(5.26) \quad \mu \left( \limsup_N |A_N(F_1, F_2, \dots, F_\ell)| > \epsilon \right) \leq \epsilon.$$

Let  $\epsilon > 0$  be given and let  $G_1$  be a linear combination of indicators so that

$$(5.27) \quad \left( \int_X |F_1 - G_1|^2 \right)^{1/2} < \delta,$$

where we'll choose  $\delta$  later appropriately for  $\epsilon$ . By Theorem D, we have

$$\limsup_N |A_N(G_1, F_2, \dots, F_\ell)(x)| = 0 \text{ for a.e. } x.$$

We then have

$$\limsup_N |A_N(F_1, F_2, \dots, F_\ell)(x)| = \limsup_N |A_N(F_1 - G_1, F_2, \dots, F_\ell)(x)| \text{ for a.e. } x.$$

Using the boundedness of the  $F_i$  by 1 we can estimate pointwise as

$$|A_N(F_1 - G_1, F_2, \dots, F_\ell)(x)| \leq B_N(F_1 - G_1)(x) + C_N(F_1 - G_1)(x).$$

We thus have

$$\int_X \limsup_N |A_N(F_1 - G_1, F_2, \dots, F_\ell)| \leq \int_X \sup_N B_N(F_1 - G_1) + \int_X \sup_N C_N(F_1 - G_1)$$

by the Cauchy-Schwarz-Bunyakovsky inequality for the integral  $\int_X$

$$\leq \left( \int_X \left( \sup_N B_N(F_1 - G_1) \right)^2 \right)^{1/2} + \left( \int_X \left( \sup_N C_N(F_1 - G_1) \right)^2 \right)^{1/2}$$

using the maximal inequalities for  $B_N$  in eq. (5.24) and for  $C_N$  in eq. (5.25)

$$\leq \left( c \int_X (F_1 - G_1)^2 \right)^{1/2} + \left( c_\omega \int_X (F_1 - G_1)^2 \right)^{1/2}$$

by the approximation in eq. (5.27)

$$\leq c_\omega \delta.$$

Using this and Markov's inequality, we finally get

$$\begin{aligned} \mu \left( \limsup_N |A_N(F_1, F_2, \dots, F_\ell)| > \epsilon \right) &\leq \frac{1}{\epsilon} \cdot \int_X \limsup_N |A_N(F_1 - G_1, F_2, \dots, F_\ell)| \\ &\leq \frac{c_\omega \delta}{\epsilon}. \end{aligned}$$

We see, we just have to choose  $\delta$  small enough so that

$$\frac{c_\omega \delta}{\epsilon} < \epsilon$$

to get eq. (5.26).

What we have accomplished so far is that our averages  $A_N(F_1, F_2, \dots, F_\ell)$  converge a.e. to 0 for all bounded  $F_1$  and indicators  $F_2, F_3, \dots, F_\ell$ . But we can repeat the argument with  $F_2$  in place of  $F_1$  to get that  $A_N(F_1, F_2, \dots, F_\ell)$  converge a.e. to 0 for all bounded  $F_1, F_2$  and indicators  $F_3, F_4, \dots, F_\ell$ . Repeating the argument  $\ell$  times, we eventually get that  $A_N(F_1, F_2, \dots, F_\ell)$  converge a.e. to 0 for bounded  $F_1, F_2, \dots, F_\ell$ .

## 6. PROOF OF THEOREM C, THE SEMIRANDOM CONVERGENCE THEOREM

Our first observation is that the a.e. convergence of our averages

$$(6.1) \quad \frac{1}{N} \sum_{n < N} T^n F_1 \cdot (x) T^{r_n} F_2(x)$$

to  $\overline{F_1} \overline{F_2}$  needs to be proved only for indicator functions  $F_1, F_2$ . This follows from a similar maximal inequality argument as we used to extend Theorem D from indicators to bounded functions.

So, unless we say otherwise, we now assume, for the rest of the proof, that the  $F_i$  are indicator functions. We will follow the general structure of the proof in [FLW12]. As in the proof of Theorem D, it is enough to prove convergence along a sparse subsequence of the  $N$ . As in that proof, for  $\sigma > 1$ , we choose the subsequence  $I_\sigma = \{N_1 < N_2 < N_3 < \dots\}$  so that  $\sum_{n < N_i} \mathbb{1}_n \approx \sigma^i$  (and hence  $\sum_{n < N_i} U_n \approx \sigma^i$ )

$$(6.2) \quad \sum_{n < N_i} \mathbb{1}_n \geq \sigma^i \text{ but } \sum_{n < N} \mathbb{1}_n < \sigma^i \text{ for } N < N_i.$$

To be exact, since  $\mathbb{1}_n \rightarrow 0$ , we have

$$\sum_{n < N_i} \mathbb{1}_n = \sigma^i + o(1) \text{ for all } i = 1, 2, 3, \dots$$

We notice, that the averages in eq. (6.1) can be written in the form

$$\frac{1}{\sum_{n < N} U_n} \sum_{n < N} U_n T^{U_1 + U_2 + \dots + U_n} F_1 T^n F_2.$$

The two main steps in proving that the above converges a.e. are that the ‘‘partial’’ expectations

$$(6.3) \quad \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} \mathbb{1}_n T^{U_1 + U_2 + \dots + U_n} F_1 T^n F_2$$

converge to  $\bar{F}_1 \bar{F}_2$ , and that the difference between the two averages

$$\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) T^{U_1+U_2+\dots+U_n} F_1 T^n F_2$$

converges to 0.

We will be able to prove the convergence of the partial expectations in eq. (6.3) for *any* sequence  $(\mathbb{1}_n)$  with  $\mathbb{1}_n \rightarrow 0$  and satisfying the dissipative property  $\sum_n \mathbb{1}_n = \infty$ . To emphasize this fact, we formulate the result explicitly

**Theorem 6.1.** *Suppose the decreasing sequence  $\{\mathbb{1}_1 \geq \mathbb{1}_3 \geq \mathbb{1}_1 \geq \dots\}$  satisfies*

$$\sum_n \mathbb{1}_n = \infty,$$

and

$$\lim_{n \rightarrow \infty} \mathbb{1}_n = 0.$$

*We then have, with probability 1, in any dynamical system,*

$$\lim_{N \rightarrow \infty} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} \mathbb{1}_n T^{U_1+U_2+\dots+U_n} F_1 T^n F_2 = \bar{F}_1 \bar{F}_2 \text{ for a.e. } x \in X \text{ and any bounded functions } F_i.$$

*Proof.* We divide the proof into two cases: first, when  $F_2$  is  $T$ -invariant, and second, when  $F_2$  is orthogonal to all  $T$ -invariant functions. The function  $F_1$  is an indicator throughout the proof.

Let thus  $F_2$  be  $T$ -invariant. We then need to prove the pointwise convergence of

$$\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} \mathbb{1}_n T^{U_1+U_2+\dots+U_n} F_1.$$

The a.e. convergence of the averages

$$(6.4) \quad \frac{1}{\sum_{n < N} U_n} \sum_{n < N} U_n T^{U_1+U_2+\dots+U_n} F_1$$

to  $\bar{F}_1$  follows from the pointwise ergodic theorem. To see this, let  $r_1 < r_2 < \dots < r_k < \dots < r_K < N$  be all the elements of our random sequence which are less than  $N$ . Since  $U_1, U_2, \dots$  is the indicator of the random sequence,  $U_n = 1$  exactly when  $n = r_k$  for some  $k \leq K$ . It follows that

$$\sum_{n < N} U_n = K,$$

and

$$U_n T^{U_1+U_2+\dots+U_n} = \begin{cases} T^k & \text{if } n = r_k \text{ for some } k \leq K, \\ 0 & \text{otherwise,} \end{cases}$$

so the average in eq. (6.4) is the  $K$ th ergodic average  $\frac{1}{K} \sum_{1 \leq k \leq K} T^k F_1$ .

Since the averages in eq. (6.4) converge a.e. to  $\bar{F}_1$ , by the strong law of large numbers, the averages

$$\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} U_n T^{U_1+U_2+\dots+U_n} F_1.$$

also converge to  $\bar{F}_1$ . We then need to show that, with full probability, the difference between the two averages converges to 0 a.e.

$$(6.5) \quad \lim_{N \rightarrow \infty} \sup_{(X, \mu, T), F_1} \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) T^{U_1+U_2+\dots+U_n} F_1 \right| = 0 \text{ with probability 1.}$$

Let  $\epsilon_N \rightarrow 0$ . Later we'll find that the fastest rate our method allows  $\epsilon_N$  to go to 0 is  $O\left((\sum_{n < N} \mathbb{1}_n)^{-1/4}\right)$  which is plenty, as we will soon see. We shall prove

$$(6.6) \quad P\left(\sup_{(X, \mu, T), F_1} \int_X \left| \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) T^{U_1+U_2+\dots+U_n} F_1 \right| > \epsilon_N\right) < \exp\left(-c\epsilon_N^2 \sum_{n < N} \mathbb{1}_n\right)$$

for all large enough  $N$ . This implies eq. (6.5) if  $\epsilon_N$  can be chosen so that, for all  $\sigma > 1$  we have

$$(6.7) \quad \sum_{N \in I_\sigma} \epsilon_N < \infty,$$

and

$$(6.8) \quad \sum_{N \in I_\sigma} \exp \left( -c\epsilon_N^2 \sum_{n < N} \mathbb{1}_n \right) < \infty,$$

where  $I_\sigma$  is defined in eq. (6.2). This is because we need to show convergence only for the subsequence  $I_\sigma$ .

Following the proof of Theorem B, a duality and a transference argument shows that eq. (6.6) follows from

$$(6.9) \quad P \left( \sup_{f \in \mathcal{F}, g \in \mathcal{G}} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \frac{1}{Q} \sum_{a < Q} g(a) f(a + (U_1 + U_2 + \dots + U_n)) > \epsilon_N \right) < \exp \left( -c\epsilon_N^2 \sum_{n < N} \mathbb{1}_n \right) \text{ for all large enough } N,$$

where  $\mathcal{G}$  is the same as before, that is, it contains  $\pm 1$  valued functions supported on the interval  $[0, Q]$

$$\mathcal{G} = \{g \mid g : [0, Q] \rightarrow \{-1, 1\}\},$$

and  $\mathcal{F}$  contains indicators supported on the interval  $[0, Q + U_1 + U_2 + \dots + U_N]$

$$\mathcal{F} = \{f \mid f : [0, Q + U_1 + U_2 + \dots + U_N] \rightarrow \{0, 1\}\}.$$

In fact, except on a set with small probability, we can replace the interval  $[0, Q + U_1 + U_2 + \dots + U_N]$  above, which depends on  $\omega$ , by the interval  $[0, Q + 2 \sum_{n < N} \mathbb{1}_n]$ . This is because when we put  $X_n = U_n - \mathbb{1}_n$  and  $t = \sum_{n < N} \mathbb{1}_n$  in Bernstein's inequality, Lemma 4.1, we get

$$(6.10) \quad P \left( U_1 + U_2 + \dots + U_N - \sum_{n < N} \mathbb{1}_n > \sum_{n < N} \mathbb{1}_n \right) < \exp \left( -c \sum_{n < N} \mathbb{1}_n \right),$$

and, of course,  $\sum_{N \in I_\sigma} \exp(-c \sum_{n < N} \mathbb{1}_n) < \infty$ . So from now on,  $\mathcal{F}$  denotes indicators on the interval  $[0, Q + 2 \sum_{n < N} \mathbb{1}_n]$

$$\mathcal{F} = \left\{ f \mid f : \left[ 0, Q + 2 \sum_{n < N} \mathbb{1}_n \right] \rightarrow \{0, 1\} \right\}.$$

The next step in our method is to divide up the interval  $[0, Q + 2 \sum_{n < N} \mathbb{1}_n]$  into residue classes  $(\text{mod } q)$  for an appropriate  $q$ . The same way as in previous proofs, we get that a good choice for  $q$  is  $q = Q$ . We also get the requirement, using the cardinality of functions supported on a single residue class  $(\text{mod } q)$  in  $[0, Q + 2 \sum_{n < N} \mathbb{1}_n]$  and the cardinality of  $\mathcal{G}$ , that

$$(6.11) \quad \frac{2 \sum_{n < N} \mathbb{1}_n}{q} + Q \leq c\epsilon_N^2 \sum_{n < N} \mathbb{1}_n.$$

Since  $q = Q$ , the left hand side of eq. (6.11) is smallest when

$$(6.12) \quad q = Q = c \left( \sum_{n < N} \mathbb{1}_n \right)^{1/2}.$$

In the above, nothing changes significantly, if we choose the constant  $c$  so that  $Q$  is an integer. Substituting eq. (6.12) into eq. (6.11), we get

$$\epsilon_N = c \left( \sum_{n < N} \mathbb{1}_n \right)^{-1/4}.$$

With these values all our requirements are satisfied including eqs. (6.7) and (6.8), and our proof of eq. (6.5) is finished.



Our next step in proving Theorem 6.1 is to consider the case when  $F_2$  is orthogonal to the invariant functions. Since in this case,  $\bar{F}_2 = 0$ , we want to prove

$$\frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} \mathbb{1}_n T^{U_1+U_2+\dots+U_n} F_1 T^n F_2 \rightarrow 0.$$

By a maximal inequality argument—similar but simpler than the one we used to extend Theorem D to bounded functions—it is enough to prove eq. (6.12) for a class of functions which generates an  $L^2$ -dense linear subspace of the orthocomplement of the  $T$ -invariant functions. This class of functions will be the “bounded” coboundaries, so those functions  $F_2$  which can be written in the form  $F_2 = TG - G$  for some bounded  $G$ .

So let  $F_2$  be a coboundary. We shall prove a much more general result: The only property of the sequence  $\{T^1 F_2(x), T^2 F_2(x), \dots\}$  we use is that  $\frac{1}{N} \sum_{L < n < L+N} T^n F_2(x)$  converges to 0 *uniformly* in  $L$  as  $N \rightarrow \infty$ . The only property of the random sequence  $\{r_1 < r_2 < r_3 < \dots\}$  we use is that it has 0 density as a consequence of the assumption that  $\mathbb{1}_n \rightarrow 0$ .

**Lemma 6.2.** *Suppose the set  $R$  of positive integers has 0 density, and let  $X(n)$  denote its indicator. Let  $\{g_1, g_2, g_3, \dots\}$  be a bounded sequence of real numbers which satisfy*

$$(6.13) \quad \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{M \leq n < M+L} g_n = 0 \quad \text{uniformly in } M.$$

*Then for any bounded sequence  $f_1, f_2, f_3, \dots$  of numbers and decreasing sequence  $w_1 \geq w_2 \geq w_3 \geq \dots$  of weights with  $\sum_n w_n = \infty$  we have*

$$\lim_{N \rightarrow \infty} \frac{1}{\sum_{n < N} w_n} \sum_{n < N} w_n f_{X(1)+X(2)+\dots+X(n)} g_n = 0.$$

*Proof.* A summation by parts argument shows that we can assume all the weights  $w_n$  are equal to 1, so we need to show

$$\frac{1}{N} \sum_{n < N} f_{X(1)+X(2)+\dots+X(n)} g_n \rightarrow 0.$$

The main observation is that the complement of  $R$  is *basically* a set which is the union of longer and longer intervals. More precisely, for any given length  $L$ , there is a  $S \subset \mathbb{R}^c$  of density 1 so that  $S$  is the union of intervals of length at least  $L$ . To see this, suppose to the contrary: there is a length  $L$  so that the set of  $k$ ,  $k = 1, 2, 3, \dots$ , for which the interval  $[kL, (k+1)L)$  contains a point from  $R$  has positive upper density, say,  $\delta > 0$ . But then the upper density of  $R$  is at least  $\delta/L$ , a contradiction.

To prove the lemma, let  $L$  be arbitrary, and consider the average

$$\frac{1}{KL} \sum_{k \leq K} \sum_{kL \leq n < (k+1)L} f_{X(1)+X(2)+\dots+X(n)} g_n.$$

Let  $k_1 < k_2 < \dots$  be a sequence so that  $R \subset \cup_i [k_i L, (k_i + 1)L)$ , and the density of  $(k_i)$  is 0. Write

$$\begin{aligned} & \frac{1}{KL} \sum_{k \leq K} \sum_{kL \leq n < (k+1)L} f_{X(1)+X(2)+\dots+X(n)} g_n \\ &= \frac{1}{KL} \sum_i \sum_{k_i L \leq n < (k_i+1)L} f_{X(1)+X(2)+\dots+X(n)} g_n + \frac{1}{KL} \sum_{k \neq k_i} \sum_{kL \leq n < (k+1)L} f_{X(1)+X(2)+\dots+X(n)} g_n. \end{aligned}$$

The first sum above goes to 0, since  $(k_i)$  has 0 density and the functions are assumed to be bounded. For the second sum, note that on each interval  $[kL, (k+1)L)$ , we have  $X(n) = 0$ , hence  $f_{X(1)+X(2)+\dots+X(n)}$  is constant. We can estimate

$$\left| \frac{1}{KL} \sum_{k \neq k_i} \sum_{kL \leq n < (k+1)L} f_{X(1)+X(2)+\dots+X(n)} g_n \right| \leq \frac{1}{K} \sum_{k \neq k_i} \left| \frac{1}{L} \sum_{kL \leq n < (k+1)L} g_n \right|.$$

By the assumption in eq. (6.13),  $\left| \frac{1}{L} \sum_{kL \leq n < (k+1)L} g_n \right|$  is uniformly small if  $L$  is big, finishing the proof of the lemma.  $\square$

This ends the proof of Theorem 6.1.  $\square$

Our last task in the proof of Theorem C is to prove that with probability 1 we have, in every dynamical system,

$$\lim_{N \rightarrow \infty} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) T^{U_1+U_2+\dots+U_n} F_1(x) T^n F_2(x) = 0 \text{ for almost every } x.$$

Let  $\delta > 0$  and

$$(6.14) \quad \epsilon_N = \left( \log \sum_{n < N} \mathbb{1}_n \right)^{-(1+\delta)}.$$

As many times before in this paper, our task is reduced to proving an inequality for the probability of averages on the integers

$$\begin{aligned} P \left( \sup_{f_i \in \mathcal{F}_i, g \in \mathcal{G}} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \frac{1}{Q} \sum_{a < Q} g(a) f_1 \left( a + (U_1 + U_2 + \dots + U_n) \right) f_2(a + n) > \epsilon_N \right) \\ \leq \exp \left( -c \epsilon_N^2 \sum_{n < N} \mathbb{1}_n \right), \text{ for all large enough } N, \end{aligned}$$

where  $Q$  is to be determined momentarily,  $\mathcal{F}_1$  contains all the indicators supported in the interval<sup>3</sup>  $[0, Q + 2 \sum_{n < N} \mathbb{1}_n]$

$$\mathcal{F}_1 = \left\{ f \mid f : \left[ 0, Q + 2 \sum_{n < N} \mathbb{1}_n \right] \rightarrow \{0, 1\} \right\},$$

$\mathcal{F}_2$  contains all the indicators supported in the interval  $[0, Q + N]$ ,

$$\mathcal{F}_2 = \left\{ f \mid f : [0, Q + N] \rightarrow \{0, 1\} \right\},$$

and  $\mathcal{G}$  contains all the  $\pm 1$  valued functions supported on the interval  $[0, Q]$

$$\mathcal{G} = \left\{ g \mid g : [0, Q] \rightarrow \{-1, 1\} \right\}.$$

We now choose the relatively prime moduli  $q_1, q_2$ , and we divide up the support of  $f_i$  according to the residue classes  $(\bmod q_i)$ . With the assumption that

$$q_1 q_2 = Q,$$

we further reduce the problem to the case when the  $f_i$  are supported on a fixed  $r_i$  residue class  $(\bmod q_i)$ . This way, following the proof of Theorem A, we get the estimate

$$(6.15) \quad P \left( \sup_{f_i \in \mathcal{F}_i, g \in \mathcal{G}} \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \frac{1}{Q} \sum_{a < Q} g(a) f_1 \left( a + (U_1 + U_2 + \dots + U_n) \right) f_2(a + n) > \epsilon_N \right) \\ \leq q_2 q_1 \exp \left( Q + \frac{\sum_{n < N} \mathbb{1}_n}{q_1} + \frac{N}{q_2} - \frac{1}{4} \epsilon_N^2 \sum_{n < N} \mathbb{1}_n \right).$$

Ignoring the negligible  $q_1 q_2 < \exp(c \log N)$  product (negligible compared to  $\exp(\epsilon_N^2 \sum_{n < N} \mathbb{1}_n)$ , which is assumed to be  $\exp(O(N^{1/2}))$ ), we see our task is to choose the parameters  $Q, q_1, q_2$  so that

$$Q + \frac{\sum_{n < N} \mathbb{1}_n}{q_1} + \frac{N}{q_2} < c \epsilon_N^2 \sum_{n < N} \mathbb{1}_n.$$

<sup>3</sup>We already showed in eq. (6.10) that we make only a small error in probability, if we replace the random interval  $[0, Q + \sum_{n < N} U_n]$  by the deterministic one  $[0, Q + 2 \sum_{n < N} \mathbb{1}_n]$ .

If  $Q$  and  $q_i$  were real variables, the left hand side is minimal if

$$\begin{aligned} Q &= c\epsilon_N^2 \sum_{n < N} \mathbb{1}_n, \\ q_1 &= c\epsilon_N^{-2}, \\ q_2 &= c \frac{N}{\epsilon_N^2 \sum_{n < N} \mathbb{1}_n}. \end{aligned}$$

We certainly can choose integer values for  $Q, q_i$  so that they are within constant multiples of the above optimal values,  $q_1$  and  $q_2$  are coprimes, and  $q_1 q_2 = Q$ . Using that  $q_1 q_2 = Q$ , we get the requirement

$$\frac{N}{\epsilon_N^4 \sum_{n < N} \mathbb{1}_n} \leq c\epsilon_N^2 \sum_{n < N} \mathbb{1}_n,$$

hence

$$N^{1/2} \leq c\epsilon_N^3 \sum_{n < N} \mathbb{1}_n,$$

which is possible, since the value of  $\epsilon_N$  in eq. (6.14) is compatible with the assumption in eq. (2.6). With this, our proof is finished.

## 7. NOTES

**7.1. Finitary version of Theorem A.** We use the setup commonly used in combinatorics: for a given  $N$ , we want to select each integer  $1, 2, \dots, N$  with a uniform probability  $\mathbb{1}_N$ . So for each  $N$ , we take  $U_{N1}, U_{N2}, \dots, U_{NN}$ , a sequence of independent, identically distributed, 0 – 1-valued random variables, with the law

$$P(U_{Nn} = 1) = 1 - P(U_{Nn} = 0) = \mathbb{1}_N, \quad n = 1, 2, \dots, N.$$

The sequence  $U_{N1}, U_{N2}, \dots, U_{NN}$  is the indicator of the set  $R_N^\omega$ ,

$$R_N^\omega = \{n | U_{Nn}(\omega) = 1\}.$$

The proof of Theorem A can easily be adjusted to give

**Theorem** (Finitary version of Theorem A). *Let  $\ell$  be a positive integer and assume*

$$\mathbb{1}_N \cdot N^{1/(\ell+1)} \rightarrow \infty.$$

*Then there is a set  $\Omega' \subset \Omega$  with  $P(\Omega') = 1$  such that for every  $\omega \in \Omega'$  the following is true: for all  $\alpha > 0$  there is  $N(\alpha)$  so that if  $A \subset [1, N]$  for some  $N > N(\alpha)$  and  $A$  has more than  $\alpha N$  elements, then  $A$  contains an arithmetic progression  $a, a + r, a + 2r, \dots, a + \ell r$  of length  $(\ell + 1)$  with some  $r \in R_N^\omega$ .*

**7.2. Our method for bounded functions.** In our paper, we chose to extend almost everywhere convergence results from indicators to bounded functions using the method of maximal inequalities. Our method can be applied directly to bounded functions, though. The modification in the proof is needed only in the definition of the sets  $\mathcal{F}$ . For example, consider the proof of Theorem D in the  $\ell = 1$  case. We need to consider the averages

$$(7.1) \quad \frac{1}{\sum_{n < N} \mathbb{1}_n} \sum_{n < N} (U_n - \mathbb{1}_n) \frac{1}{Q} \sum_{a < Q} f(a + n)$$

for  $f : [0, Q + N] \rightarrow [-1, 1]$ . As a first step, we want to replace  $f$  by a function which take on values from a sufficiently dense *discrete subset* of the interval  $[0, 1]$ . Assume, for notational simplicity, that  $\epsilon_N$  is the reciprocal of a positive integer. A dense enough set is the set

$$D = \{-1, -(\epsilon_N^{-1} - 1)\epsilon_N, \dots, -2\epsilon_N, -\epsilon_N, 0, \epsilon_N, 2\epsilon_N, \dots, (\epsilon_N^{-1} - 1)\epsilon_N, 1\},$$

so we define

$$\mathcal{F} = \{f | f : [0, Q + N] \rightarrow D\}.$$

Now, if we replace  $f$  in eq. (7.1) by a function from  $\mathcal{F}$ , then with probability very close to 1, we only make a  $2\epsilon_N$  error. It follows that it is enough to consider functions from  $\mathcal{F}$  in eq. (7.1). The cardinality of  $\mathcal{F}$  is not much bigger than it was before: since  $|D| = c\epsilon_N^{-1}$ , we have

$$|\mathcal{F}| < |D|^{2N} < \exp(cN \log \epsilon_N^{-1}).$$

Since in Theorem D,  $\epsilon_N$  is a negative power of  $\log N$ , this extra  $\log \epsilon_N^{-1}$  factor would account for an extra  $\log \log N$  speed—negligible as the speed in Theorem D is stated, but it would not be negligible if we wanted to state the best possible speed our proof could give. For example, it's clear from the proof that we can weaken the assumption in eq. (5.21) of Theorem D to

$$\mathbb{U}_n \cdot \frac{n^{1/2}}{(\log n)^2 \log \log^{2+\delta} n} \rightarrow \infty \text{ for some } \delta > 0.$$

But if we used our method directly to bounded functions instead of maximal inequalities we would have needed the stronger

$$\mathbb{U}_n \cdot \frac{n^{1/2}}{(\log n)^2 \log \log^{3+\delta} n} \rightarrow \infty \text{ for some } \delta > 0$$

assumption.

This is the reason why we decided to use the well-known maximal inequality techniques, and hence we avoided the slight inefficiency and complications in our method when dealing with bounded functions instead of indicators.

**7.3. Convexity methods.** We can also use the more geometric *convexity* methods to extend almost sure convergence results from indicators to bounded functions. This method is signified by the Krein-Milman theorem, though we need only the simplest version of it for finite sets. Let us briefly give here the representation needed to do the extension from indicators to nonnegative functions bounded by 1<sup>4</sup>. Let

$$\begin{aligned} \Gamma &= \{g|g : [1, K] \rightarrow \{0, 1\}\} = \{0, 1\}^K, \\ \mathcal{F} &= \{f|f : [1, K] \rightarrow [0, 1]\} = [0, 1]^K. \end{aligned}$$

We want to show that for each  $f \in \mathcal{F}$  there is a probability measure  $\mu_f$  on  $\Gamma$  so that

$$(7.2) \quad f(b) = \int_{\Gamma} g(b) d\mu_f(g).$$

The requirement of the representation uniquely determines the measure. For any  $b$ , eq. (7.2) looks like

$$f(b) = 1 \cdot \mu_f\{g|g(b) = 1\}.$$

It follows

$$\begin{aligned} \mu_f\{g|g(b) = 1\} &= f(b), \\ \mu_f\{g|g(b) = 0\} &= 1 - f(b). \end{aligned}$$

Since the sets  $\mu\{g|g(b) = 1\}$  and  $\mu\{g|g(b) = 0\}$ ,  $b = 1, 2, \dots, K$  are all the cylinder sets of  $\Gamma = \{0, 1\}^K$ , the measure  $\mu_f$  is uniquely determined as a product measure.

With this representation, it's a simple exercise to extend, say, eq. (5.8), from indicators to nonnegative functions bounded by 1.

There is only a slight complication when we have to extend results for multiple averages from indicators to bounded functions. For example, for double averages, we need to be able to represent the product  $f_1(b_1)f_2(b_2)$  in an integral form. This can be done using the representation in eq. (7.2) for  $f_1$  and  $f_2$  separately and Fubini's theorem

$$\begin{aligned} f_1(b_1)f_2(b_2) &= \left( \int_{\mathcal{G}} g(b_1) d\mu_{f_1}(g) \right) \left( \int_{\mathcal{G}} g(b_2) d\mu_{f_2}(g) \right) \\ &= \int_{\mathcal{G} \times \mathcal{G}} g_1(b_1)g_2(b_2) d(\mu_{f_1} \times \mu_{f_2})(g_1, g_2), \end{aligned}$$

which is just a fancy (but useful) way of writing

$$f_1(b_1)f_2(b_2) = \mu_{f_1}\{g|g(b_1) = 1\} \cdot \mu_{f_2}\{g|g(b_2) = 1\}.$$

<sup>4</sup>Similar representation applies for extending results for  $\pm 1$ -valued functions to functions bounded by 1.

**7.4. Extension to commuting transformations.** In short: our method in its present form doesn't work for commuting transformations. For example, we have

**Unsolved problem 7.1.** Let  $b < 1/(\ell + 1)$  and  $u_n = n^{-b}$ .

Show that with probability 1 in any dynamical system, for any commuting transformations  $T_1, T_2, \dots, T_\ell$ , the averages

$$\frac{1}{\sum_{n < N} u_n} \sum_{n < N} U_n \cdot T_1^n F_1 \cdot T_2^n F_2 \cdots T_\ell^n F_\ell$$

converge in  $L^1$  norm.

The good news is that the extension of Theorem C to commuting transformations is true. In other words, the semirandom averages

$$\frac{1}{N} \sum_{n < N} T_1^n F_1(x) \cdot T_2^{r_n} F_2(x)$$

do converge almost everywhere for commuting measure preserving transformations  $T_i$  as long as the random sequence  $\{r_1 < r_2 < r_3 < \dots\}$  satisfies  $r_n/n^{2-\epsilon}$  for some positive  $\epsilon$ . The proof of this will be discussed in an upcoming paper.

#### APPENDIX A. STRONG LAW OF LARGE NUMBERS

Here we give a quick proof of the strong law of large numbers in a form needed in our paper. The result is due to Kolmogorov.

**Theorem A.1** (Strong law of large numbers for uniformly bounded rv's). *Suppose the random variables  $X_1, X_2, \dots$  are non-negative, uniformly bounded*

$$0 \leq X_n \leq c,$$

*(pairwise) uncorrelated*

$$(A.1) \quad \mathbb{E} X_n X_m = \mathbb{E} X_n \cdot \mathbb{E} X_m, \text{ for } n \neq m,$$

*and dissipative*

$$\sum_N \mathbb{E} X_n = \infty.$$

*Then*

$$\lim_{N \rightarrow \infty} \frac{1}{\sum_{n < N} \mathbb{E} X_n} \sum_{n < N} X_n = 1 \text{ with probability 1.}$$

For an example of two random variables which are nonnegative, uncorrelated but not independent, take  $X_1(x) = 1 + \sin(2\pi x)$  and  $X_2(x) = 1 + \sin(2\pi 2x)$  defined on the unit interval  $[0, 1]$ . (Looking at the graph of the two functions, by inspection we can find a  $\lambda$ ,  $0 < \lambda < 2$ , so that the level sets  $\{X_1 > \lambda\}$  and  $\{X_2 > \lambda\}$  are nonempty, but disjoint.)

*Proof.* We want to use a subsequence argument. Define the set  $I = \{N_1 < N_2 < \dots < N_i < \dots\}$  of indices by

$$\sum_{n < N_i} \mathbb{E} X_n \geq i^2 \text{ but } \sum_{n < N} \mathbb{E} X_n < i^2 \text{ for } N < N_i.$$

In words:  $N_i$  is the smallest index  $N$  for which  $\sum_{n < N} \mathbb{E} X_n \geq i^2$ . That  $I$  is well defined and infinite follows from the dissipative assumption  $\sum_N \mathbb{E} X_n = \infty$ . Since the  $X_n$  are uniformly bounded, we have

$$(A.2) \quad \sum_{n < N_i} \mathbb{E} X_n = i^2 + O(1).$$

We first want to show that we have a.e. convergence along the subsequence  $I$

$$(A.3) \quad \lim_{i \rightarrow \infty} \frac{1}{\sum_{n < N_i} \mathbb{E} X_n} \sum_{n < N_i} X_n = 1 \text{ with probability 1.}$$

We show this in the form

$$\sum_i \left( \frac{1}{\sum_{n < N_i} \mathbb{E} X_n} \sum_{n < N_i} X_n - \mathbb{E} X_n \right)^2 < \infty \text{ with probability 1.}$$

This follows if the above series is integrable, so from

$$(A.4) \quad \mathbb{E} \sum_i \left( \frac{1}{\sum_{n < N_i} \mathbb{E} X_n} \sum_{n < N_i} X_n - \mathbb{E} X_n \right)^2 = \sum_i \mathbb{E} \left( \frac{1}{\sum_{n < N_i} \mathbb{E} X_n} \sum_{n < N_i} X_n - \mathbb{E} X_n \right)^2 < \infty.$$

A consequence of the multiplicativity assumption in eq. (A.1) is orthogonality

$$\mathbb{E}(X_n - \mathbb{E} X_n)(X_m - \mathbb{E} X_m) = 0 \text{ for } n \neq m,$$

and hence we have

$$\mathbb{E} \left( \sum_{n < N_i} X_n - \mathbb{E} X_n \right)^2 = \sum_{n < N_i} \mathbb{E}(X_n - \mathbb{E} X_n)^2.$$

Since  $\mathbb{E}(X_n - \mathbb{E} X_n)^2 \leq \mathbb{E} X_n^2 \leq c \mathbb{E} X_n$ , we have

$$\mathbb{E} \left( \frac{1}{\sum_{n < N_i} \mathbb{E} X_n} \sum_{n < N_i} X_n - \mathbb{E} X_n \right)^2 \leq \frac{c}{\sum_{n < N_i} \mathbb{E} X_n},$$

which, by eq. (A.2), establishes eq. (A.4).

We now want to show that the  $N$ th average is close to the  $N_i$ th if  $N_i \leq N < N_{i+1}$ . We can estimate, since the  $X_n$  are non-negative,

$$\begin{aligned} \frac{1}{\sum_{n < N} \mathbb{E} X_n} \sum_{n < N} X_n &\leq \frac{1}{\sum_{n < N_i} \mathbb{E} X_n} \sum_{n < N_{i+1}} X_n \\ &= \frac{\sum_{n < N_{i+1}} \mathbb{E} X_n}{\sum_{n < N_i} \mathbb{E} X_n} \cdot \frac{1}{\sum_{n < N_{i+1}} \mathbb{E} X_n} \sum_{n < N_{i+1}} X_n. \end{aligned}$$

This implies, by eq. (A.3) and eq. (A.2),

$$\begin{aligned} \limsup_{N \rightarrow \infty} \frac{1}{\sum_{n < N} \mathbb{E} X_n} \sum_{n < N} X_n &\leq \lim_{i \rightarrow \infty} \frac{(i+1)^2 + O(1)}{i^2 + O(1)} \\ &= 1. \end{aligned}$$

Similar estimate shows that

$$\liminf_{N \rightarrow \infty} \frac{1}{\sum_{n < N} \mathbb{E} X_n} \sum_{n < N} X_n \geq 1,$$

finishing our proof. □

## REFERENCES

- [BL96] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996. 2
- [Bos83] M. Boshernitzan. Homogeneously distributed sequences and Poincaré sequences of integers of sublacunary growth. *Monatsh. Math.*, 96(3):173–181, 1983. 2, 3, 4
- [Bou87] J. Bourgain. Ruzsa's problem on sets of recurrence. *Israel J. Math.*, 59(2):150–166, 1987. 3
- [Bou88] J. Bourgain. On the maximal ergodic theorem for certain subsets of the integers. *Israel J. Math.*, 61(1):39–72, 1988. 4, 6, 20
- [Bou90] J. Bourgain. Double recurrence and almost sure convergence. *J. Reine Angew. Math.*, 404:140–161, 1990. 5, 19
- [BP98] J. Brüdern and A. Perelli. Goldbach numbers in sparse sequences. *Ann. Inst. Fourier (Grenoble)*, 48(2):353–378, 1998. 4
- [CG10] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. *ArXiv e-prints*, November 2010, 1011.4310. 8
- [Chr11] M. Christ. On random multilinear operator inequalities. *ArXiv e-prints*, August 2011, 1108.5655. 3, 5
- [CL84] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, 112(2):143–175, 1984. 4
- [DL96] J.-M. Derrien and E. Lesigne. Un théorème ergodique polynomial ponctuel pour les endomorphismes exacts et les  $K$ -systèmes. *Ann. Inst. H. Poincaré Probab. Statist.*, 32(6):765–778, 1996. 19
- [ES77] P. Erdős and A. Sárközy. On differences and sums of integers. II. *Bull. Soc. Math. Grèce*, 18(2):204–233, 1977. 3
- [ES78] P. Erdős and A. Sárközy. On differences and sums of integers. I. *J. Number Theory*, 10(4):430–450, 1978. 2, 3
- [FHK07] N. Frantzikinakis, B. Host, and B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. *J. Reine Angew. Math.*, 611:131–144, 2007. 2, 4
- [FLW06] N. Frantzikinakis, E. Lesigne, and M. Wierdl. Sets of  $k$ -recurrence but not  $(k+1)$ -recurrence. *Ann. Inst. Fourier (Grenoble)*, 56(4):839–849, 2006. 2
- [FLW12] N. Frantzikinakis, E. Lesigne, and M. Wierdl. Random sequences and pointwise convergence of multiple ergodic averages. *Indiana Univ. Math. J.*, 61:585–617, 2012. 3, 5, 6, 21

- [Fra10] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *J. Anal. Math.*, 112:79–135, 2010. 4
- [Fur77] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977. 2, 4, 9, 13, 14
- [FW09] N. Frantzikinakis and M. Wierdl. A Hardy field extension of Szemerédi’s theorem. *Adv. Math.*, 222(1):1–43, 2009. 2
- [HK01] B. Host and B. Kra. Convergence of Conze-Lesigne averages. *Ergodic Theory Dynam. Systems*, 21(2):493–509, 2001. 4
- [HK05a] B. Host and B. Kra. Convergence of polynomial ergodic averages. *Israel J. Math.*, 149:1–19, 2005. Probability in mathematics. 4
- [HK05b] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005. 4, 16
- [JLW99] R. L. Jones, M. Lacey, and M. Wierdl. Integer sequences with big gaps and the pointwise ergodic theorem. *Ergodic Theory Dynam. Systems*, 19(5):1295–1308, 1999. 5, 6, 20
- [Kar71] A. A. Karacuba. Estimates of trigonometric sums by the method of I. M. Vinogradov, and their applications. *Trudy Mat. Inst. Steklov.*, 112:241–255, 388. (errata insert), 1971. Collection of articles dedicated to Academician Ivan Matveevič Vinogradov on his eightieth birthday, I. 4
- [Lei05] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Israel J. Math.*, 146:303–315, 2005. 4
- [Les93] E. Lesigne. Équations fonctionnelles, couplages de produits gauches et théorèmes ergodiques pour mesures diagonales. *Bull. Soc. Math. France*, 121(3):315–351, 1993. 19
- [Pet75] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York-Heidelberg, 1975. Translated from the Russian by A. A. Brown, *Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82*. 9
- [Sár78a] A. Sárközy. On difference sets of sequences of integers. I. *Acta Math. Acad. Sci. Hungar.*, 31(1–2):125–149, 1978. 2
- [Sár78b] A. Sárközy. On difference sets of sequences of integers. III. *Acta Math. Acad. Sci. Hungar.*, 31(3–4):355–386, 1978. 2
- [Sze75a] E. Szemerédi. On sets of integers containing no  $k$  elements in arithmetic progression. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 2, pages 503–505. Canad. Math. Congress, Montreal, Que., 1975. 2
- [Sze75b] E. Szemerédi. On sets of integers containing no  $k$  elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. 2
- [Var59] P. Varnavides. On certain sets of positive density. *J. London Math. Soc.*, 34:358–360, 1959. 9, 13
- [Wie89] M. Wierdl. *Almost everywhere convergence and recurrence along subsequences in ergodic theory*. ProQuest LLC, Ann Arbor, MI, 1989. Thesis (Ph.D.)—The Ohio State University. 2, 4
- [WZ12] T. D. Wooley and T. D. Ziegler. Multiple recurrence and convergence along the primes. *Amer. J. Math.*, 134(6):1705–1732, 2012. 2, 4
- [Zie07] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20(1):53–97 (electronic), 2007. 16

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CRETE, HERAKLION, GREECE

UNIVERSITÉ FRANÇOIS RABELAIS, CNRS, LMPT UMR 7350, TOURS, FRANCE

DEPARTMENT OF MATHEMATICAL SCIENCES, THE UNIVERSITY OF MEMPHIS, MEMPHIS, TN, USA